

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocollaboration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:41:57 ; Search time 225.02 Seconds  
(without alignments)  
10613.880 Million cell updates/sec

Title: US-10-611-442-2\_COPY\_17000\_17596

Perfect score: 597  
Sequence: 1 aacacgagcgtgtgacgc.....caacgagcgcagcg99gcg 597

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 9267905 seqs, 2000278028 residues

Total number of hits satisfying chosen parameters: 18535810

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA New:\*

- 1: /SIDS5/ptodata/2/pubpna/US08 NEW PUB. seq:\*
- 2: /SIDS5/ptodata/2/pubpna/US06 NEW PUB. seq:\*
- 3: /SIDS5/ptodata/2/pubpna/US07 NEW PUB. seq:\*
- 4: /SIDS5/ptodata/2/pubpna/PCT\_NEW PUB. seq:\*
- 5: /SIDS5/ptodata/2/pubpna/US09 NEW PUB. seq:\*
- 6: /SIDS5/ptodata/2/pubpna/US09 NEW PUB. seq:\*
- 7: /SIDS5/ptodata/2/pubpna/US10 NEW PUB. seq:\*
- 8: /SIDS5/ptodata/2/pubpna/US10 NEW PUB. seq:\*
- 9: /SIDS5/ptodata/2/pubpna/US10 NEW PUB. seq:\*
- 10: /SIDS5/ptodata/2/pubpna/US10 NEW PUB. seq:\*
- 11: /SIDS5/ptodata/2/pubpna/US11 NEW PUB. seq:\*
- 12: /SIDS5/ptodata/2/pubpna/US11 NEW PUB. seq:\*
- 13: /SIDS5/ptodata/2/pubpna/US11 NEW PUB. seq:\*
- 14: /SIDS5/ptodata/2/pubpna/US11 NEW PUB. seq:\*
- 15: /SIDS5/ptodata/2/pubpna/US60\_NEW\_PUB. seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	48.4	8.1	1685	11	US-11-096-568A-22628
C 2	46.8	7.8	1092	8	US-10-467-657-4105
C 3	43.2	7.2	1773	11	US-11-096-568A-22025
C 4	42	7.0	1986	11	US-11-096-568A-17491
C 5	41.8	7.0	1447	11	US-11-096-568A-26890
C 6	41.8	7.0	1444	11	US-11-096-568A-20988
C 7	41.8	7.0	2506	11	US-11-096-568A-19863
C 8	41.2	6.9	1492	11	US-11-096-568A-10536
C 9	40.4	6.8	11070	14	US-11-075-185-34
C 10	40.4	6.8	78869	14	US-11-075-185-1
C 11	40	6.7	13440	11	US-11-096-568A-19408
C 12	39.6	6.6	1284	9	US-10-714-887-121
C 13	39.4	6.6	1416	8	US-10-858-730-195
C 14	39.4	6.6	1694369	7	US-10-506-454-1890
C 15	39.2	6.6	908	11	US-11-096-568A-5433
C 16	39	6.5	1591	11	US-11-096-568A-5433
C 17	39	6.5	88421	14	US-11-205-109-1
C 18	38.2	6.4	1216	8	US-10-821-234-838

C 19	38	6.4	849	11	US-11-096-568A-21222	Sequence 21222, A
C 20	38	6.4	3408	8	US-10-858-730-40	Sequence 40, Appl
C 21	38	6.3	3513	8	US-10-858-730-142	Sequence 142, Appl
C 22	37.8	6.3	1095	8	US-10-793-626-3075	Sequence 3075, Ap
C 23	37.8	6.3	1435	11	US-11-096-568A-11892	Sequence 11892, A
C 24	37.8	6.3	3017	8	US-10-793-626-4422	Sequence 4422, Ap
C 25	37.8	6.3	4185	8	US-10-793-626-3646	Sequence 3646, Ap
C 26	37.8	6.3	5301	14	US-11-075-185-40	Sequence 40, Appl
C 27	37.8	6.3	78869	14	US-11-075-185-1	Sequence 1, Appl1
C 28	37.6	6.3	1022	11	US-11-096-568A-10781	Sequence 10781, A
C 29	37.6	6.3	2124	11	US-11-096-568A-23989	Sequence 23989, A
C 30	37.4	6.3	1187	11	US-11-096-568A-22600	Sequence 22600, A
C 31	37.4	6.3	1629	9	US-10-501-035-167	Sequence 167, App
C 32	37.4	6.3	35521	9	US-10-714-887-117	Sequence 242, App
C 33	37.2	6.2	1096	9	US-10-714-887-117	Sequence 117, App
C 34	37.2	6.2	1212	14	US-11-000-688-635	Sequence 635, App
C 35	37.2	6.2	1328	11	US-11-096-568A-8975	Sequence 8975, Ap
C 36	37.2	6.2	3342	14	US-11-169-041-95	Sequence 95, Appl
C 37	37.2	6.2	157224	14	US-11-112-908-51	Sequence 51, Appl
C 38	37.2	6.2	161726	14	US-11-112-908-48	Sequence 48, Appl
C 39	37.2	6.2	161726	14	US-11-112-908-52	Sequence 52, Appl
C 40	37.2	6.2	170189	14	US-11-112-908-50	Sequence 50, Appl
C 41	37	6.2	765	9	US-10-714-887-125	Sequence 125, Appl
C 42	37	6.2	1296	11	US-11-096-568A-23788	Sequence 23788, A
C 43	36.8	6.2	1247	11	US-11-096-568A-10198	Sequence 10198, A
C 44	36.8	6.2	1264	11	US-11-096-568A-23146	Sequence 23146, A
C 45	36.8	6.2	2078	14	US-11-024-959-52	Sequence 52, Appl

#### ALIGNMENTS

RESULT 1  
US-11-096-568A-22628/c  
Sequence 22628, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nickolai et al.  
TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptide  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 22628  
LENGTH: 1685  
TYPE: DNA  
ORGANISM: zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(1685)  
OTHER INFORMATION: Ceres Seq. ID no. 12409422  
US-11-096-568A-22628

Query Match	8.1%	Score 48.4	DB 11	Length 1685
Best Local Similarity	47.0%	Pred. No. 0.00058		
Matches 185	Conservative 0	Mismatches 206	Indels 3	Gaps 1
QY	23	GGCCGAGCGCGACACCTGCTCAGATCTGCGCCGAGTGTGAGGCGGT	82	
DB	933	GGCCGCGTGTGAGATGCTCCCTTGAATGACCCCGGAGCCCGGTGCGCCG	874	
QY	83	CGCGGCGCGACGCGGAGAACTGGACCGGAGCGCACCCCGCTGAGACGACG	142	
DB	873	CCGACCGCGACCGGACCGGCGGACGCGACGATGCTGTGTAAGAGGAGGAA	814	
QY	143	CGGTGAGCGCGTGTGAGGAGGATCCGACCGCTTCAATGATTCGCTGACG	202	
DB	813	AACGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	754	
QY	203	CGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	262	
DB	753	ACGTATGAGCGCGCGTGTGCGGCGCGCGGTGTGCGGCGGTGCGGCGG	694	

QY 263 AGAGTGTGAGTGTGCGCCGAGTTCAGCGTGTGGCGAGTCTTTGGCGTGGCGAT 322  
DB 693 CCACGGCGCGCGCGCCGAGGTGTAGCCGAGCCCGGCGCCCATATAGGTGAGTCTTGG 634  
QY 323 AAAGCTTGTGAGATGTGCGCTG---CTTGCCTTGGCGCGCGCGCTCCGCGAGCT 379  
DB 633 GAGCTTCACCGTCTTGTGTGTGTACCGGCGAGCGGAGACCTTGAAGCGCGAGA 574  
QY 380 CGAGAAACGACAGGTGGAAGCCGAGCGCGGAC 413  
DB 573 CGGACCCCGCGGCTCTGCGCCGTAAGCGCGAC 540

## RESULT 2

US-10-467-657-4105/c  
Sequence 4105, Application US/10467657  
Publication No. US20050260581A1  
GENERAL INFORMATION:  
APPLICANT: CHIRON SPA  
APPLICANT: PONTANA Maria Rita  
APPLICANT: PIZZA Mariagrazia  
APPLICANT: MASIGNANI Vega  
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/467,657  
CURRENT FILING DATE: 2003-08-11  
PRIOR APPLICATION NUMBER: GB-0103424.8  
PRIOR FILING DATE: 2001-02-12  
NUMBER OF SEQ ID NOS: 9218  
SOFTWARE: SeqMln9, version 1.04  
SEQ ID NO 4105  
LENGTH: 1092  
TYPE: DNA  
ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-4105

Query Match 7.8%; Score 46.8; DB 8; Length 1092;  
Best Local Similarity 53.2%; Pred. No. 0.0016;  
Matches 99; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 36 CACCTGTGACGATCTGCGCCGAGGAGTGTGAGGCGCGTCCGCGCCGAGC 95  
DB 414 CACTGTCCGATGATTTGGCGAGCGGTTAAATGGGTTGAAGCCCTGCGCGCGTGA 355  
DY 96 GCAGAACTGGCACGGAAAGCGCACCCCGCTGAGACGAGCGCGGTGTAGCCGTC 155  
DB 354 ACAAAATGTACATTCGAAAGCGCACCTTGCGAGGAAATGCGAGAGCGTCCGCGATC 295  
QY 156 GTGGGAGAGATCCGACCGTGTGATGAAATTTGCGTTCGACCACTGGAACGAACTT 215  
DB 294 CGATTGCGGAGATGAAGAGGTTTCCAGCGCGTTCGCCGTACGACATCAAAAGCAATT 235  
QY 216 TGTCTG 221  
DB 234 TCGAGT 229

## RESULT 3

US-11-096-568A-22025  
Sequence 22025, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 22025  
LENGTH: 1773

TYPE: DNA  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(1773)  
OTHER INFORMATION: Ceres Seq. ID no. 12407294  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1742)-(1742)  
OTHER INFORMATION: n is a, c, t, g, unknown, or other  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1743)-(1743)  
OTHER INFORMATION: n is a, c, t, g, unknown, or other  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1767)-(1767)  
OTHER INFORMATION: n is a, c, t, g, unknown, or other  
US-11-096-568A-22025

Query Match 7.2%; Score 43.2; DB 11; Length 1773;  
Best Local Similarity 54.0%; Pred. No. 0.017;  
Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 241 CTGCGCTGCGAGAGGTGACGAGAGGTGTCGAGTTCGCCGTGAGCTTCAAGCTGTG 300  
DB 668 CCGCGCGCGCGCGGACCGGTGGCGGCGTCCGCTCCGCGCTCCGCGCG 727  
QY 301 GCCGAGTCTTGGCGCTGCGCATTAAGCTTTCGAAAGTGTGCGGTGCTTCCGTCGCG 360  
DB 728 TCCGCGAGCTGCGCTGCTTCACTTCGACCTTCGAGAGAGAGCCCTTACGCTTCG 787  
QY 361 CCGAGCGCTCCGCGAGCTCGAGAGAGGTGCAAGAC 401  
DB 788 CGAGCGCTTCGAGAGCTGTCGCGCGCTGAGAGCGGCG 828

## RESULT 4

US-11-096-568A-17491/c  
Sequence 17491, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 17491  
LENGTH: 1986  
TYPE: DNA  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(1986)  
OTHER INFORMATION: Ceres Seq. ID no. 12358713  
US-11-096-568A-17491

Query Match 7.0%; Score 42; DB 11; Length 1986;  
Best Local Similarity 46.6%; Pred. No. 0.038;  
Matches 135; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 5 CGAGCTTTGACCGCGCGCGCGCGAGACCTGTCGACGATCGCCCGGTAGCA 64  
DB 934 CGATGATTTGACGACCTGCGCGAGCGGATGATCGAGTGTGTAAGCGGCG 875  
QY 65 GGTGTGCTGAGCGCGTGTGCGCGCGCGAGAACTGCGACGCGAGCGCACCCCG 124  
DB 874 CCGCGCGCGGAGCGCGCGCGCGAGCGCGTGTGAGGAGGAGGAGGAGGAG 815  
QY 125 CTTGAGAGAGAGCGAGCGGCTGTAGCCGTGTGCGAGGATCCGACCGTTCGATGA 184



Db 814 CGAGAGCGCCCGAAGCCGATCCAGCGGGTACGGTGAACCTTGTCTTCGCGCTTGAGAGA 755  
QY 185 AATTGCGGTGACAGAGTGCAGACAGAACTTTGTGCTGCGTTCCTCCCTGCGACTGC 244  
Db 754 ACTTGTGGCGCGGAAGTTGATCCCAATGCGGTGATGAGGGGAGCAAGATTAAGTGGCGA 695  
QY 245 GCTCGCGGAGGGTGCAGAGAGGTGCTGAGTTCGCCGTGATGCTTTCAG 294  
Db 694 ACCTGGCGGACCTCCCGCGGACGTCCGCGCTCTGGCGGATGAGACAGAGAC 645

## RESULT 5

US-11-096-568A-26890/c  
; Sequence 26890, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 26890  
; LENGTH: 1147  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(1147)  
; OTHER INFORMATION: Ceres Seq. ID no. 13628192  
US-11-096-568A-26890

Query Match 7.0%; Score 41.8; DB 11; Length 1147;  
Best Local Similarity 45.7%; Pred. No. 0.042;  
Matches 185; Conservative 0; Mismatches 217; Indels 3; Gaps 1;

QY 1 AACACAGGCTGTGACGCGCGGCGGAGCGCCAGACCTGCTGCGACATCTGCCCGGT 60  
Db 676 AACTCATACATGAGAGAGTGGTGAAGCGACTGTCGAGATGATGTAACCTGC 617  
QY 61 ACACAGGTGTGCTTGAAGCGCGTGCAGCGCCGACGCGCAGAACTGACACCGAAGCGCAC 120  
Db 616 CGGCGCGCGGATGAGAGCGTGCAGAGCGGCGGAGACGTGGGGCGGTGCGTGAAG 557  
QY 121 CCGGCTGAGACGAGCGAGGCGGTGAGCGCTGCTGCGACGATCCGACCGTCTCG 180  
Db 556 ACCTGGTCCGCAATTGAACGTGGGCGACGACGCGCTTGAGACCTGTTGCGATGAGGGG 497  
QY 181 ATGAATATGCGCTCG--ACCACTGGAACAGAACTTTGTGCTGCTGCCCTGCTG 237  
Db 496 ACACACTTGTGCTGCTGACTGAGTCCGAGGAGAGTGAAGTGTGGGAGAGTGTGAGACG 437  
QY 238 CGACTGCGCTGCGCGAGGAGTGCAGAGAGGTGCTGAGTTCGCCGTGATGCTTCAAGCTG 297  
Db 436 TCGGGCGGAGAGAGAGAGAGAGCGAGCTGAGATTTGACATCTGCAAGTCTTGGTCCCGAG 377  
QY 298 TGGGCGGAGTCTTGTGGCTGCCGATTAAGCTTGTGCAAGATGTGCGCTGCTGCTGCG 357  
Db 376 TTGAGAGAGAAAGTGTGGCGGCGCGCGGATGTGCAAGATGAAGGGCTTTCGACGACG 317  
QY 358 CGGCGGAGCGGCTCCGCGGAGTCCGAGAACTGAGAACTGAGTCCAGAACCC 402  
Db 316 GGCAAGAGAAATGGGTGCCCTGCGCGAGGCTCTCCGAGGACAC 272

## RESULT 6

US-11-096-568A-20988/c  
; Sequence 20988, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; TITLE OF INVENTION: Theby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20988  
; LENGTH: 1444  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(1444)  
; OTHER INFORMATION: Ceres Seq. ID no. 12391561  
US-11-096-568A-20988

Query Match 7.0%; Score 41.8; DB 11; Length 1444;  
Best Local Similarity 43.5%; Pred. No. 0.042;  
Matches 190; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

QY 15 GAGCGCGGCGCGAGCGCCAGACCTGCTGCAAGATCTGCCCGGTGAGAGAGTGTGCTT 74  
Db 1107 GAGCGCGGTGAGAGTCCGCGGTGAGTGTGAACTGAAAGCCAGGTGAGGTGCAGAG 1048  
QY 75 GAGCGCGTGCAGCGCCGACGCGCAGAACTGACACCGAAGCGCACCCGCTGAGACGA 134  
Db 1047 GCTGTGCGCGCGCGCTTGTCCGTGAGCGGTGATCCGCTGTCTCTGCGGTATGGGAC 988  
QY 135 GAGCAGCGCGGTGAGCGGTGCTGCGACGAGATCCGACCGTCTGATGAATATTCGCTC 194  
Db 987 GCGGTGGCGGAGATGGCGAAGACCGCGTCCAGCTTCAACATGACGTTGTGACGGCGTTC 928  
QY 195 GAGCAGCTGAGAACAGAACTTTGTGCTGCGCTTCCCTGCTGCGTCCGCTGCGCGAG 254  
Db 927 CATGCGCGTGAAGAGAGCGCGGAGAGGCGCTTGAACGCGCCGCGCTTTCGACGCG 868  
QY 255 GGTGACAGAGAGGTGCTGAGTTCGCCGTGATGCTTCAAGTGTGGCCGAGTCTTTCG 314  
Db 867 GTGACGTGAGAGCGGCTGCGGCTCCAGAGTGAAGTGTGATGTGCTCTCTCTGCTGCA 808  
QY 315 CTGCGATTAAGCTTGTGAGATGTGCGCTGCTGCTGCGCGCGACGCGCTCCGC 374  
Db 807 CTCGCGCGCTGCGGCGCGCGACGATGAGAGCGGTGAGCGCGCGCGCGCTGCTGCTG 748  
QY 375 GAGCTCGGAGAACGACAGTGCAGAACCGACCGCGGAGTCTGCTGCGCGAATGGG 434  
Db 747 GACGAAGTGAACCCCGACGCGCTGACCCACGTGAAGTGCAGTGAAGGCGCGGTTTGG 688  
QY 435 TAGACCAACGACCTGGG 451  
Db 687 GTTCCGATGAAGTGG 671

## RESULT 7

US-11-096-568A-19863/c  
; Sequence 19863, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 19863  
; LENGTH: 2506  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(2506)  
; OTHER INFORMATION: Ceres Seq. ID no. 12375134  
US-11-096-568A-19863

Query Match	7.0%;	Score 41.8;	DB 11;	length 2506;
Best Local Similarity	46.2%;	Pred. No. 0.043;		
Matches 139; Conservative	0;	Mismatches 162;	Indels 0;	Gaps 0;

Oy	3	CACAGAGCTGTGTGAAGCCGGAGCCCGAGGGCCAGAGACCTGTGCACAGATCTTGCCCGGGTAG	62
Db	1360	CTCGAACCTGGTGTGATCTTGGAGAAACATGAGAGGGCTCCCGCTCCGGGTGCGCGCTGTGCAG	1901
Oy	63	CAGGTGTGCTTTGAGGCGCGGTGCGCGGCCGACGCGCGCAAACTGGACACCGGAAGGCGCACCC	122
Db	1900	GAGCGTGTGTCAACAGACAGGGCGGCGAGCTTCTGCATCAGCGGCAAGGCTGGGCGTGGCCGC	1841
Oy	123	CGCTGAGACAGACGACGAGCGGTGTAGCCGTCTGTGCGACAGATCCGCACCGTCTGAT	182
Db	1840	GTCCTCCGCGTACGTGAACACAGGACTCCGGGTGCATGGCGCTGCATCAGCTCTTCTCGCT	1781
Oy	183	GAATTTGCCGTGCACCAAGCTGGAACAGAACTTTGTGCTGTGGCTTCCCTGTGTGCAGCT	242
Db	1780	GAAGCGAGCGCTGAGAGACTCGCCCGAAGGGTTCAATGTCTACAGCACTTCTTGGCCACTG	1721
Oy	243	GCGCTCGGCGAGGGTGCAGAGAGGTGCTGCAGTTGCCCGTAGTGTCTTACGCGTGTGGGC	302
Db	1720	GGTCAACGGTGTCTTTCACCGACGCGCTTATGTTCTCTCGAGGTGGCGCAGGTCCACGCG	1661
Oy	303	C 303	
Db	1660	C 1660	

```

RESULT 8
US-11-096-568A-10536/c
; Sequence 10536, Application US/11096568A
; Publication NO. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theryd
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 10536
; LENGTH: 1492
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1492)
; OTHER INFORMATION: Cereas Seq. ID no. 13596376
US-11-096-568A-10536

```

Query Match	Similarity	Score	DB	Length
Beet Local	47.3%	Pred. No. 0	063	
Matches	124	Conservative	0	Mismatches 138; Indels 0; Gaps 0
Qy	6	GAGCCTGTGACGCCCGGCGCGAGCGCAGCACTGTGCGACGATCTCGCCCGTACGAG	65	
Db	918	GAGCTTGTGTGCGCGCGCGCCCGCTGTGTGTGATGACAGCAGCAGGAGCGCTTACGCGGCAC	859	
Qy	66	GTGTGCTTTGAGGCGGTGCGGCGCCGAGCGCGAGAGACTGTGGCAGCGAAGCGCCACCCGCG	125	
Db	858	GCGCGCCCGCTTCACTTTATGTGGCGTCAACGCGCGATGAAGTACTCGTTGCACTTGTCTT	799	
Qy	126	CTGAGACGAGACGCGAGCGCGGTGAGCGCGTGTGTCGACGAGATCCGACCGTCTCGATGAA	185	
Db	798	CTCGCGGAGAGCGCGCGCGGTGTCTCACTGTGTGACGACACCGGGGTGTAGAGAGCGCA	739	
Qy	186	ATTGCGCGTGAACGAGCTGAGAGAGAACTTTGTGCTGTGCTTCCCTGTGTCGACTGCG	245	
Db	738	CTTGGAGAGGTCCACCCCGGGCTGGAAACCGGTATGGCGCGTGTGCGCAAGACACGCGCCC	679	
Qy	246	CTCGCGGAGCGGTGACGAGAGG	267	

Db 678 GGCCGGGAGGTCGAAGTGAGG 657

```

RESULT 9
US-11-075-185-34
; Sequence 34, Application US/11075185
; Publication No. US2005026643A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, PALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 34
; LENGTH: 11070
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-34

```

Query Match	6.8%	Score 40.4	DB 14	Length 11070
Best Local Similarity	44.5%	Pred. No. 0.11		
Matches 161, Conservative	0	Mismatches	201	Indels 0; Gaps 0

QY	10	CTGTTGAAGCCGAGGCGCCGAGGCGCAGACCTGCTGACAGATCTGCGCCGCTAGCAGGTGT	69
Db	8126	CGGCGCGCGCGCGCGCGCGCGCGAGCGCGCATCTGTACCGGCTGCGCTGAGACGACGTGG	8185
QY	70	CGCTTGAAGCCGCTGCGCGGCCGAGCGCGGACAAGTGGACAAGGAAAGCGCACCCGCTGA	129
Db	8186	CGCTTGAAGACACACGCGCTTTGTGCGCGAGAGACAGCTGTGTGTGTGGCGGACGCGCGC	8245
QY	130	GACGAGACGCGAGCGGTTGATCGCTGTGTGGGAGACGGAATCGGCACCGCTTGGATGAATTTG	189
Db	8246	TGGCGCGCGCGCTGCGGTGACACGCGGAGTGGCGGCGCGCGCGAGCTGCGCGTCCGCTGC	8305
QY	190	CGGTGACACAGCTCGACAGGAATTTTGTCTGTGACTTCCCTGATGTCGACTGCTCG	249
Db	8306	CGGACGCGCGCGCGCGCGCGCACCGCGCTGTGTGTGACCTGACCGCGGACGCGCGCGCGC	8355
QY	250	GCGAGGCTGCACGAGAGGTCTGTGAGTTGCCCGTAGTCTTACGAGTGTGGGCGGAGTCT	309
Db	8366	CGGTGTGTGACGCGCTGTGACGCGCGCGCGCGCGACGCGCTGTCCCTGTGTGAGGGGTGTGC	8425
QY	310	TTGCGCTCGCCGATTAAGCTTGTTCGAAGATGTGCGTGTGCGCGCGCGCGACGCGC	369
Db	8426	TGGCGCGCGCGCGAGCTGGCGGCGACGAGAGCTGTGTGTGTGACGCGCGCGCGGTTGGCG	8485
QY	370	TC 371	
Db	8486	TC 8487	

RESULT 10  
 US-11-075-185-1  
 ; Sequence 1, Application US/11075185  
 ; Publication No. US20050266434A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: REEVES, CHRISTOPHER D  
 ; APPLICANT: JULIEN, BRYAN  
 ; APPLICANT: REID, RALPH  
 ; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS  
 ; FILE REFERENCE: 010099\_03  
 ; CURRENT APPLICATION NUMBER: US/11/075,185  
 ; CURRENT FILING DATE: 2005-03-07

```

RESULT 10
US-11-075-185-1
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099_03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07

```

PRIOR APPLICATION NUMBER: US 60/551,103  
PRIOR FILING DATE: 2004-03-08  
PRIOR APPLICATION NUMBER: US 60/568,290  
PRIOR FILING DATE: 2004-05-04  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 1  
LENGTH: 78869  
TYPE: DNA  
ORGANISM: Sorangium cellulosum  
US-11-075-185-1

Query Match 6.8%; Score 40.4; DB 14; Length 78869;  
Best Local Similarity 44.5%; Pred. No. 0.12;  
Matches 161; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 10 CTGTTGACCGCGCGCCCGGAGCGGCGGACGCTGTCACGATCTGCGCCGCTGACGAGTGT 69  
DB 12327 CGGCGCGCGCGCGCGCGCGCGGAGCGGATCTGTCACCGCTGCGCTGACGAGCTGG 12386  
QY 70 CGCTTAGGCGCGTCCGCGCGCGGACGCGGACGATGCGGCGGAGCGGCGGCGCTGTA 129  
DB 12387 CGGTGACGACGACGCGCGCTTTGTGCGGAGGACGCTGCTGCGCGGCGGAGCGGCG 12446  
QY 130 GACGAGACGCGGCGCGGTGACCGCTGCGGCGGACGATCCGCGCTCTGATGAAATTG 189  
DB 12447 TGGCGGCGCGCGCTCGGTGACGCGCGGCGGCGGCGGCGGCGGCGGCGCTGCGTGC 12506  
QY 190 CGCTGACGACGCTGCGGACGAACTTTGTGCTGCGCTTCCCTGCGTGCAGTCCGCTCG 249  
DB 12507 CGGACGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12566  
QY 250 GCGAGGCGTGCACGAGGCGTGTGAGTTCGCGGATGCGCTTCAAGCGTGGGCGGAGTCT 309  
DB 12567 CGGTCTGACGCGCGCTGTCACGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 12626  
QY 310 TTGCGCTGCGCGATTAAGCTTTGTCGAGATGTCGCGTTCGCGTTCGCGCGGCGGCG 369  
DB 12627 TGGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12686  
QY 370 TC 371  
DB 12687 TC 12688

RESULT 11  
US-11-096-568A-19408/C  
Sequence 19408, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 19408  
LENGTH: 1340  
TYPE: DNA  
ORGANISM: Zea mays subsp. mays  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(1340)  
OTHER INFORMATION: Ceres Seq. ID no. 12373073  
US-11-096-568A-19408

Query Match 6.7%; Score 40; DB 11; Length 1340;  
Best Local Similarity 46.4%; Pred. No. 0.14; Mismatches 150; Indels 0; Gaps 0;  
Matches 130; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
QY 3 CAGGAGCTGTGACGCGCGCGCGCGGAGCGGCGGACGCTGTCGACGATCTCGCGCGGTAG 62

DB 1089 CAGCGCCCAATTAGCCACGGAAGCTTAAGCAGAGCGGCGCCACGCTCAGCGCTGGGGA 1030  
QY 63 CAGGTGCTGTTAGGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 122  
DB 1029 GCTACTAATGACACGCTTTGTGTGTTGCGCATGAAGTACACGCGCGCGCGCG 970  
QY 123 CGCTGAGACGAGACCGGCGGCTGAGCGCTGTCGCGGCGGCGGCGGCGGCGGCGGCT 182  
DB 969 CGACGAGGCGGCGGCGGCGGCTGCTGATCGCTGACGACCCCAAGCTCTCTGCGCGCTGCC 910  
QY 183 GAAATTGCGGTGACGAGCTGGAACAGAACTTTGCTGTGCGCTTCCCTGTCGACT 242  
DB 909 GTACATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 850  
QY 243 GCGCTGCGGAGGCTGACGAGGCTGTCGAGTTCGCG 282  
DB 849 GGGGCGGCGCTGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 810

RESULT 12  
US-10-714-887-121/C  
Sequence 121, Application US/10714887  
Publication No. US20060015972A1  
GENERAL INFORMATION:  
APPLICANT: Mendel Biotechnology, Inc.  
APPLICANT: HEARD, Jacqueline  
APPLICANT: RIECHMANN, Josef Luis  
APPLICANT: CREELMAN, Robert  
APPLICANT: RATCLIFFE, Oliver  
APPLICANT: CANALES, Roger  
APPLICANT: REPERTI, Peter  
APPLICANT: KUMIMOTO, Roderick W  
APPLICANT: GUTTERSON, Neal  
APPLICANT: REUBER, T. Lynne  
APPLICANT: PINEDA, Omaira  
APPLICANT: SHERMAN, Bradley K  
TITLE OF INVENTION: PLANT TRANSCRIPTIONAL REGULATORS OF DROUGHT STRESS  
FILE REFERENCE: MB10058-CIP  
CURRENT APPLICATION NUMBER: US/10/714,887  
PRIOR FILING DATE: 2003-11-13  
PRIOR APPLICATION NUMBER: 10/412,699  
PRIOR FILING DATE: 2003-04-10  
PRIOR APPLICATION NUMBER: 09/506,720  
PRIOR FILING DATE: 2000-02-17  
PRIOR APPLICATION NUMBER: 60/135,134  
PRIOR FILING DATE: 1999-05-20  
PRIOR APPLICATION NUMBER: 09/394,519  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: 09/533,392  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 09/533,029  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 09/532,591  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 09/533,030  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 60/125,814  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: 09/713,994  
PRIOR FILING DATE: 2000-11-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 430  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 121  
LENGTH: 1284  
TYPE: DNA  
ORGANISM: Oryza sativa (japonica cultivar-group)  
FEATURE:  
OTHER INFORMATION: G3690 Predicted polypeptide sequence is orthologous to G2999  
US-10-714-887-121

Query Match 6.6%; Score 39.6; DB 9; Length 1284;  
Best Local Similarity 49.1%; Pred. No. 0.18;

Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```
QY 18 GCCGGGCGCCGAGCGCACTGCTGCAAGATCGCCCGGTAGCAGGTGTGCTTGG 77
DB 675 GGGGGGCGAGAGCGAGCGAGCGCGCGCGCGCGCTCGCGCGGTGGAAGCT 616
QY 78 GCCGTCCGCGCCGAGCGCGAGAACTGCGACGCGAGAGCGCACCCCGCTGAGACGAGAC 137
DB 615 ACCGTGCGACCGCGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCAT 556
QY 138 GCAAGCGGTGTAGCGCTGCTGCGAGCGAGTCGCAACGCTTCGATGAATTCGCTGAC 197
DB 555 GAACTCCCGCGAGCGCTGCGAGCAGTCGCGCGCCCATCCGCGCGCGGTCTTGAAGCA 496
QY 198 CAGCTGAGAGAGAACTTGTGCTGCTGCTTCC 231
DB 495 CTCCCTTACTCTCAGCTCGCGCGCGCGCC 462
```

## RESULT 13

```
US-10-858-730-195/C
; Sequence 195, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 195
; LENGTH: 1416
; TYPE: DNA
; ORGANISM: Thermobifida fusca
US-10-858-730-195
```

Query Match 6.6%; Score 39.4; DB 8; Length 1416;  
Best Local Similarity 47.4%; Pred. No. 0.2;  
Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
QY 165 GATCGGCAACGCTTCGATGAATTCGCTGCAACGCTCGAAGCACTTGTGCTG 224
DB 1134 GGTCAACGCGCGGTGGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1075
QY 225 GCTTCCCTGTCGACATGCGCTCGGAGGGTTCAGACGAGTGTGATTTGCCGTA 284
DB 1074 ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
QY 285 GTGCTTCAAGCGTGTGGGCGGAGTCTTTCGCTGCGGATTAAGCTTGTGGAAGTGC 344
DB 1014 TTGACACGACGAGTATTCACGCTGACCGGCTCAAGGTTCGGGCGGCTGATGATGAC 955
QY 345 TGCTTGGCGTTCGCGCGCGCGCGCGCGCTCGGAGCTCGAAGACGACAGTGAAGAC 404
DB 954 GCGGTACATGCGGTTCGAGTGTCTGACGACATGGGCGAGGTTCGAACAGTGTACACCA 895
QY 405 CGGCGCGAC 413
```

DB 894 GGCACCGCG 886

## RESULT 14

```
US-10-506-454-1690/C
; Sequence 1690, Application US/10506454
; Publication No. US20060068386A1
; GENERAL INFORMATION:
; APPLICANT: Slesarev, Alexi I
; APPLICANT: Mezhevaya, Katia V
; APPLICANT: Polushin, Nikolai N
; APPLICANT: Shcherbina, Olga V
; APPLICANT: Shakhova, Vera V
; APPLICANT: Malykh, Andrei G
; APPLICANT: Kozavkin, Sergei A
; TITLE OF INVENTION: The Complete Genome and Protein Sequences of the Hyperthermophilic
; TITLE OF INVENTION: Methanopyrus Kandleri AV19 and Monophyly of Archaeal Methanogens
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: FID001
; CURRENT APPLICATION NUMBER: US/10/506,454
; PRIOR FILING DATE: 2004-08-31
; PRIOR APPLICATION NUMBER: PCT/US03/06664
; PRIOR FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: 60/361,742
; NUMBER OF SEQ ID NOS: 1722
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1690
; LENGTH: 1694969
; TYPE: DNA
; ORGANISM: Methanopyrus kandleri
US-10-506-454-1690
```

Query Match 6.6%; Score 39.4; DB 7; Length 1694969;  
Best Local Similarity 57.9%; Pred. No. 0.25; Indels 51; Gaps 0;  
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```
QY 282 GTAGTGTTCAGCGGTGGGCGCGAGTCTTTCGCTGCGGATTAAGCTTGTGAGATGTC 341
DB 11340 GTGAGGCGCATCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11281
QY 342 GCGTCTTTCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401
DB 11280 GACCACCGGAACTCCAGTCTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 11221
QY 402 C 402
DB 11220 C 11220
```

## RESULT 15

```
US-11-096-568A-5433/C
; Sequence 5433, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; PRIOR FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 5433
; LENGTH: 908
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(908)
; OTHER INFORMATION: Ceres Seq. ID no. 14309488
US-11-096-568A-5433
```

Query Match 6.6%; Score 39.2; DB 11; Length 908;

Best Local Similarity 54.1%; Pred. No. 0.23;  
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

QY      78  GCCGTGCGGCGCCGAGCGCGAGAACTGGCAACGCGAAGCGCACCCCGCTGAGACGAGAC 137
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      531  GGGGAGAGAGCACC CGCGGACGACGCGGCAAGAGAGTGGAGCGCAGCCAGCGCTCGAC 472
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      138  GCAGCGCGTGTAGCCGTCGTGGCGAGCGATCCGCAACCGTCTCGATGAATTGCGCTGAC 197
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      471  GCAGCGCGTGTGAAACCGCTGGCTGCACTGCGGAGCACC CGCATGTCTGGCCGTCCTC 412
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      198  CAGCTCGAACAGGAACTTTGTCGCTTG 225
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      411  GAACTCGGCGAGGAGATGGCGACACTCG 384
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
  
```

Search completed: April 7, 2006, 03:50:18  
Job time : 229.02 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:15 ; Search time 2299.02 seconds  
(without alignments)  
12149.485 Million cell updates/sec

Title: US-10-611-442-2\_COPY\_17000\_17596

Perfect score: 597

Sequence: 1 aacacgagcctgtgacgc.....cacgaacgcgcagcg99cgc 597

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gest1:\*  
10: gb\_gest2:\*  
11: gb\_gest3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61.2	10.3	188	7	CN026633 UMC-p4civ
2	56.4	9.4	773	10	CL685281 PRI01404
3	56	9.4	883	7	CK151489 FGAS03405
4	55.2	9.2	561	9	BH899909 OES00976
5	54.4	9.1	872	7	CK152875 FGAS03595
6	54.4	9.1	925	10	CNS0091P
7	54.4	9.1	1133	8	DR741095 FGAS00102
8	54.4	9.1	717	3	BI955767 HVSMM002
9	53.8	9.0	655	8	CV948578 PVrpb_12
10	53.8	9.0	671	8	CV948570 PVrpb_12
11	53.8	9.0	849	7	BR265573 HV CR001
12	53.4	8.9	857	7	CK152637 FGAS03571
13	52.8	8.8	339	2	BE415712 MML038.G0
14	52.8	8.8	528	6	CD922535 G750.1037
15	52.8	8.8	771	7	CK124607 BBS182410
16	52	8.7	517	3	BM322439 P1C1_4.G1
17	52	8.7	532	6	CF432914 NT1_19.F
18	52	8.7	541	1	AM924660 NS1_71.H0
19	52	8.7	602	7	CN151959 WOUNDI_78
20	52	8.7	622	6	CF432927 NT1_19.G
21	52	8.7	635	7	CN150225 WOUNDI_67
22	52	8.7	705	6	CF429132 PH1_19.H0

C 23	52	8.7	712	7	CN142245	CN142245 WOUNDI_4
C 24	52	8.7	743	7	CN140012	CN140012 OX1_33_BO
C 25	52	8.7	765	3	CD432831	CD432831 ETRI_33.G
C 26	51.4	8.6	925	10	CNS0091P	AL053013 Drosophila
C 27	51	8.5	351	7	CN026960	CN026960 UMC-p4civ
C 28	51	8.5	357	7	CN027260	CN027260 UMC-p4civ
C 29	50.4	8.4	648	6	CD233094	CD233094 SSI_11.H1
C 30	50.4	8.4	657	6	CD233402	CD233402 SSI_13.CO
C 31	50.4	8.4	663	6	CD233416	CD233416 SSI_13.BO
C 32	50.2	8.4	627	6	CA085330	CA085330 SCILAM209
C 33	49.4	8.3	348	7	CN026914	CN026914 UMC-p4civ
C 34	48.4	8.1	642	10	CG381651	CG381651 OGVD78TH
C 35	48.4	8.1	710	9	BZ529592	BZ529592 OGJUR36TM
C 36	48.4	8.1	720	9	BZ529589	BZ529589 OGJUR36TC
C 37	48.4	8.1	794	9	CC668550	CC668550 OGJUR36TV
C 38	48.2	8.1	581	6	CB213001	CB213001 OMT03281
C 39	47.2	7.9	669	8	DR800406	DR800406 ZM_BFB002
C 40	47.2	7.9	677	8	DR800783	DR800783 ZM_BFB003
C 41	47.2	7.9	680	6	CA195481	CA195481 SCE2S109
C 42	47.2	7.9	681	9	CC678957	CC678957 SCWHG922TH
C 43	47.2	7.9	771	7	CO093637	CO093637 GR_BA15F
C 44	47.2	7.9	862	2	BG310173	BG310173 HVSME001
C 45	47.2	7.9	870	9	CC637331	CC637331 OGCX32TH

#### ALIGNMENTS

RESULT 1  
LOCUS CN026633 188 bp mRNA linear EST 09-SEP-2004  
DEFINITION UMC-p4civ1-006-e01 4-Cell-Embryo-(in vivo) p4civ Sus scrofa cDNA  
3', mRNA sequence.

ACCESSION CN026633  
VERSION CN026633.1 GI:48721169  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 188)  
AUTHORS Whitworth,K., Springer,G.K., Forrester,L.J., Spollen,W.G., Ries,J., Lamberson,W.R., Bivens,N., Murphy,C.N., Mathaiagan,N., Green,J.A. and Prather,R.S.

TITLE Developmental expression of 2489 gene clusters during pig embryogenesis: an expressed sequence tag project  
JOURNAL Biol. Reprod. 71 (4), 1230-1243 (2004)  
PUBMED 15175238

COMMENT Contact: DNA Core Facility (Swine Project)

Animal Science - RS Prather  
University of Missouri-Columbia  
M616 Medical Sciences Bldg., Columbia, MO 65212, USA  
Tel: (573)882-0428  
Fax: (573)884-5552

FEATUES  
source Email: porcine@net.missouri.edu  
POLYA=yes.

Location/Qualifiers  
1..188

/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"

/clone\_lib="p4civ"  
/note="Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source:

Slaughterhouse-derived oocytes were collected, and after cumulus cell removal were used for germinal vesicle stage oocytes, or were matured in vitro (with cumulus cells attached), in vitro fertilized and cultured. In vivo produced 4-cell and blastocyst stage embryos were collected on days 3 and 6, respectively. Zona Pelliculae were removed from the embryos prior to mRNA isolation. Expanded descriptions of how the tissues were collected



can be found at the following URL:  
<http://genome.mnet.missouri.edu/Syline/Methods.html>.

Library Construction (PCR Protocol): The amount of mRNA that was recovered from oocytes and embryos was quite limiting and was not sufficient for library production with a standard protocol. Therefore, PCR-based protocol was utilized for producing libraries. Poly-A RNA was isolated by using the MicroPoly(A) Pure kit from Ambion (cat. #1918). The mRNA was reverse transcribed with a NotI-tag-dT18 oligonucleotide and a SMART oligonucleotide (Clontech) modified to contain a SalI site to generate full-length cDNA with a sequence complementary to the SMART oligonucleotide. Sequences within the SMART and dT oligonucleotides were used as primers to amplify the cDNAs by PCR with pfu turbo polymerase (Stratagene). The resulting PCR products were purified, digested with NotI and SalI and size fractionated by using a Chroma Spin-400 followed by a Chroma Spin-1000 column (Clontech). Purified cDNA from each PCR reaction was ligated into the pCMV-SPORT6 vector. Preliminary Library Characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (-3' 96-well plates) to confirm library quality (e.g. the presence of short polyA tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.) and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. Bioinformatics work was performed by GK Springer's bioinformatics group (WG Spoelen, JS Ries, A Guillen) in Computer Science and Health Management and Informatics Departments at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core Facility at: [porcine@net.missouri.edu](mailto:porcine@net.missouri.edu). TAG TISSUB=4-Cell-Embryo-(Invivo) TAG\_SEQ=GTGCCC"

## ORIGIN

Query Match 10.3%; Score 61.2; DB 7; Length 188;  
 Best Local Similarity 61.1%; Pred. No. 6.7e-05;  
 Matches 99; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

37 ACCTGCTGCAGATCTGCGCCGGTACGAGTGTGCTTGAAGCCGTGCGGCCGAGCGG 96  
 Db AGCTGACCGATGATCTCCGAGTGTCTAAATTGCGGAGAAACCTGCTTCCCGGTCAA 85  
 97 CAGAACTGGACGCGGAAAGCGCACCCGCTGAGACGACGACGCGGTGTAGCCGTG 156  
 Db CAGAAACGGGAGTTGACGCGACACCCGCTGAGACGATGACCAACGTGCCGCGGTC 145  
 157 TGGGACGAGATCCGACCGTCTCGATGAATTGCGGTGACG 198  
 Db TTTTCGGGATGTAGACCGTCTCGACGCGATTCCTGCCCC 187

RESULT 2 773 bp DNA linear GSS 09-JUL-2004  
 CL685281  
 LOCUS PRI0140d.C01.2 - PRI0140d.BR (773) Mixed stage fosmid library of P.  
 pacificus var. California Pristionchus pacificus genomic, genomic  
 survey sequence.  
 ACCESSION CL685281 GI:50193428  
 VERSION CL685281.1  
 KEYWORDS GSS.  
 SOURCE Pristionchus pacificus  
 ORGANISM Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.  
 REFERENCE 1 (bases 1 to 773)  
 AUTHORS Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.  
 TITLE Appadri: an Acedb database for the nematode satellite organism

JOURNAL Pristionchus pacificus  
 PUBMED Nucleic Acids Res. 32 (1), D421-D422 (2004)  
 14681447  
 COMMENT Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@tuebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end  
 sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.  
 Location/Qualifiers  
 1. 773  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pBf10s-5 Fosmid vector"

## FEATURES

## ORIGIN

Query Match 9.4%; Score 56.4; DB 10; Length 773;  
 Best Local Similarity 59.3%; Pred. No. 0.0012;  
 Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

36 CACCTGCTGCAGATCTGCGCCGGTACGAGTGTGCTTGAAGCCGTGCGGCCGAGCG 95  
 Db CACTGACCGCATTAATTTCCACACCCGAGGTTCGGTTAAAGCCCTGTGGCGGTGA 440  
 96 GCAGAACTGGACGCGGAAAGCGCACCCGCTGAGACGACGACGCGGTGTAGCCGTG 155  
 Db 441 ACAGAAATTTACATCCGACGCGACACCCGCTGAGAAAGACGACGACGCGTGC 500  
 156 GTGGCGACGATCCGACCGCTCTCGATGAATTGCGGTGAC 197  
 Db 501 GTCTTCGGGATATACACCGTTTGCACGCGGTATGCCAAC 542

RESULT 3 883 bp mRNA linear EST 05-DEC-2003  
 CK151489/c  
 LOCUS PRAS034056 Triticum aestivum FGAS: Talc3 Triticum aestivum cDNA,  
 DEFINITION mRNA sequence.  
 ACCESSION CK151489  
 VERSION CK151489.1 GI:38969425  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 1 (bases 1 to 883)  
 REFERENCE Allard,F., Crosby,M.L., Danyluk,J., Eudes,F., Frick,M., Gaudet,D.,  
 Gensewein,B., Graf,R., Gulick,P., Hyman,L.D., Laroche,A.,  
 Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilsson,D.,  
 Penniket,C., Roach,J.D. and Sarhan,F.  
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
 Unpublished (2003)  
 COMMENT Contact: Mm L Crosby  
 Bioinformatics  
 University of Saskatchewan, Department of Computer Science  
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 Saskatchewan, S7N 5A9, Canada  
 Tel: 306 966 1769  
 Fax: 306 966 2033  
 Email: Egas\_este@cs.usask.ca  
 This sequence is the direct result of the Base calling software  
 phred (default parameters). It is the raw base calls. To aid in the  
 identification of the high quality insert the software Lucy

(default parameters) has been run on this sequence. Lucy identified the region [100,563].  
Plate: Talt316 row: e column: 09.  
Location/Qualifiers

## FEATURES

source

1..883  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="wheat line CI 14106"  
/db\_xref="taxon:4565"  
/lab\_host="DH5 alpha"  
/clone\_lib="Triticum aestivum FGAS: Talt3"  
/note="Organ: Crown; Vector: pGEN-T; SSH (suppression subtractive hybridization) cDNA library from genotype CI14106 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 H) (driver). Nitro-pyrole anchored oligo-dT priming and non-directional cloning."

## ORIGIN

Query Match 9.4%; Score 56; DB 7; Length 883;  
Best Local Similarity 58.6%; Pred. No. 0.0015; Mismatches 75; Indels 12; Gaps 1;  
Matches 123; Conservative 0;

QY 24 CCCGAGCGCCGAGCCTGTGTCAGATCTGCCCGGTAGCGGTGTGCTTGAGGCGGTC 83  
DB 427 CTCTATGGCCAAACCTGCTCGACATCTGTGCCCTTAAGTTCTCGCGAAGCTTCC 368  
QY 84 GCGGCCGAGCGCCGAGAACTGCGACCGGAGGCGCACCCGCTTGAGACGACGACGC 143  
DB 367 CTTCGCGATGCGCACAGAACGACGAGGGGCGACCGACCTGTGACGAAACACAGGC 308  
QY 144 GGTGTAGCC-----GTGCTGGCGACGAGATCCGACCTGTCTGATGAAATGCC 191  
DB 307 GGTGACCTTAGGGGTGCCCTGTGTCGCAACGCGGATCCCACTGTCCACCAACTGTT 248  
QY 192 GTCCAGCAGCTCGAACAGAACTTGTGCT 221  
DB 247 GTCTCTCACTTGAGAAAGTATCTTGTTGT 218

RESULT 4  
BH899909/c 561 bp DNA linear GSS 30-AUG-2002  
LOCUS Ots00976 *Ostreococcus tauri* genomic shotgun library *Ostreococcus*  
DEFINITION tauri genomic clone ota04c05.g 5', genomic survey sequence.  
ACCESSION BH899909  
VERSION BH899909.1 GI:22551404  
KEYWORDS GSS.  
SOURCE *Ostreococcus tauri*  
ORGANISM *Ostreococcus tauri*  
Bukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Mamiellales;  
Mamiellaceae; *Ostreococcus*.  
1 (bases 1 to 561)  
Derelle, E., Ferraz, C., Lagoda, P., Eychenne, S., Cooke, R., Regad, F.,  
Saban, X., Courties, C., Delaeny, M., Demaille, J., Picard, A. and  
Moreau, H.  
DNA libraries for sequencing the genome of *Ostreococcus tauri*  
(Chlorophytae, Prasinophyceae): the smallest free-living eukaryotic  
cell  
J. Phycol. 38 (6), 1150-1156 (2002)  
Contact: Moreau H  
Laboratoire Arago  
CNRS UMR 7628  
BP 44, Avenue Fontaule, 66651 Banyuls sur mer, France  
Tel: (33) 468887309  
Fax: (33) 468887398  
Email: h.moreau@obs-banyuls.fr  
Seq primer: reverse  
Classe: shotgun.

TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

Location/Qualifiers  
1..561  
/organism="Ostreococcus tauri"

/mol\_type="genomic DNA"  
/strain="OTTH0595"  
/db\_xref="taxon:70448"  
/clone\_lib="ota04c05.g"  
/note="Vector: Bluescript, Site 1: EcoRV, Site 2: EcoRV;  
Shotgun library prepared after sonication of the genomic  
DNA. Blunt ligation in EcoRV site of Bluescript. Size  
selection of the inserts after agarose electrophoresis  
between 1 and 3 Kb."

## ORIGIN

Query Match 9.2%; Score 55.2; DB 9; Length 561;  
Best Local Similarity 65.3%; Pred. No. 0.0024; Mismatches 43; Indels 0; Gaps 0;  
Matches 81; Conservative 0;

QY 24 CCCGAGCGCCGAGCCTGTGTCAGATCTGCCCGGTAGCGGTGTGCTTGAGGCGGTC 83  
DB 529 CTCAACTCTAGACCTTGCGACAGATCTCACTGCCGAGAGGTTCCGCTTGAAGCTTG 470  
QY 84 GCGGCCGAGCGCCGAGAACTGCGACCGGAGGCGCACCCGCTTGAGACGACGACGC 143  
DB 469 CAGTTCAGTGGCGCAAAACGAACTTCACTGCGGATCTTACCTGCGATGACACAGCGC 410  
QY 144 GGTG 147  
DB 409 CGTG 406

RESULT 5  
CK152875/c 872 bp mRNA linear EST 05-DEC-2003  
LOCUS FGAS035952 *Triticum aestivum* FGAS: Talt3 *Triticum aestivum* cDNA,  
DEFINITION mRNA sequence.  
ACCESSION CK152875  
VERSION CK152875.1 GI:38972365  
KEYWORDS EST.  
SOURCE *Triticum aestivum* (bread wheat)  
ORGANISM *Triticum aestivum*  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; *Triticum*.  
1 (bases 1 to 872)  
Allard, F., Crosby, W.L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,  
Genswein, B., Gatz, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
Link, M.G., McCarthy, E.L., Monroy, A., Murak, I., Nilson, D.,  
Peniket, C., Roach, J.L. and Sahau, F.  
Functional Genomics of Abiotic Stress In Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: Egae\_esc@cs.usask.ca  
This sequence is the direct result of the Base calling software  
Phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [124,739].  
Plate: Talt351 row: A column: 23.  
Location/Qualifiers  
1..872  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="wheat line CI 14106"  
/db\_xref="taxon:4565"  
/lab\_host="DH5 alpha"  
/clone\_lib="Triticum aestivum FGAS: Talt3"  
/note="Organ: Crown; Vector: pGEN-T; SSH (suppression  
subtractive hybridization) cDNA library from genotype

## FEATURES

source

Location/Qualifiers  
1..872  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="wheat line CI 14106"  
/db\_xref="taxon:4565"  
/lab\_host="DH5 alpha"  
/clone\_lib="Triticum aestivum FGAS: Talt3"  
/note="Organ: Crown; Vector: pGEN-T; SSH (suppression  
subtractive hybridization) cDNA library from genotype

## ORIGIN

CT14106 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, heated and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 h) (driver). Nitro-pyrole anchored oligo-dT priming and non-directional cloning."

Query Match 9.1%, Score 54.4; DB 7; Length 872;  
Best Local Similarity 58.1%; Pred. No. 0.0039;  
Matches 122; Conservative 0; Mismatches 76; Indels 12; Gaps 1;

QY 24 CCCGAGCGCCAGCACTGCTGACGATCTGCCCCGCTAGACAGTGTGCTTGAAGCCGTC 83  
DB 451 CTCTATGGCCCAAGACTGCTGCAATCTGTCGCCCTTAAAGTTCTCCGCGAAACCTCC 392  
QY 84 GCGGCGCCGACGCGGAGAACTGCGACGCGGACCCCGCTGAGACGAGACGACGC 143  
DB 391 CTCCCACTGCGACAGAAAGCAAGCAGCGAGGAGCGGACCTGTGACGAAACAGAGC 332  
QY 144 GGTGTAGCC-----GTGCTGCGACGAGATCCGACCGCTCTGATGAATTGCG 191  
DB 331 GGTAGCTTACGCGCGTGTCTGTCAACGGGATCCCAACGCTTCCACCACTGTT 272  
QY 192 GTGACCGAGCTCGAAGCAGAACTTTGTCT 221  
DB 271 GTCCCTCCCACTTGAGAGTATCTTGTTGT 242

RESULT 6  
CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACR19016 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL053013  
VERSION AL053013.1 GI:4934461  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)  
AUTHORS Genoscope.  
TITLES Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage; BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

## COMMENT

- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCT-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Source  
1..925  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone\_lib="BACR19016"  
/clone\_lib="RPCT-98"  
/note="end : TET3"

## ORIGIN

Query Match 9.1%, Score 54.4; DB 10; Length 925;  
Best Local Similarity 11.4%; Pred. No. 0.004;  
Matches 42; Conservative 180; Mismatches 148; Indels 0; Gaps 0;

QY 9 CTTGTTAGCCCGGCGCCGACGCGGACCTGCTGCAATCTCCGCGGTAGACGTC 68  
DB 555 SSGSGYKGGSSGSSGSCSCSSSCSCSCBCCCCGSCSYCCSSBSKCSSTS 614  
QY 69 TCGCTTAGGCGCCGCGCCGACGCGGACGAGAACTGCGACGAGAGCGACCGGCTG 128  
DB 615 BSGSCSSSSKSVGCTGSCSSSSSCSSSSSTSSSTSTSTSKSSSGSSSSSYTTSKS 674  
QY 129 AGACGAGACGACGCGGTGTAGCCGCTGTGCGACGAGATCCGACCGCTTCATGAATT 188  
DB 675 TSAAGSGSMAGGSGSTGTSSTSSSSSVSSGSKSTBSGGBSSSGSSSS 734  
QY 189 GCGGTGACCAAGCTCCAGAACTTTGTGCTGTGCTTCCCTGTGAGACTGCGCTC 248  
DB 735 STSBSBSTSTSSSSSSSVSSSTGCTCCCSYSSSTSSSTSMGTSBSGSSSVGT 794  
QY 249 GCGAGGCTGACGAGAGTCTGCAAGTCCCGTACGCTTACGCGCTGTGCGCGAGTC 308  
DB 795 SSSSDSTGSCSCCCCTCTCCSTTBMBCYTSTGCGSSSSGKGVTKCGCGGSSSTN 854  
QY 309 TTTCGCTGCCGATTAAGCTTGTGAAAGATGTCGCTGCTTCCGCTTCCCGCGACGC 368  
DB 855 GMBGTSSACSSSSSSSSSVSSSKSSAASSSVSSSGSSGVSNSASKSSGSSVS 914  
QY 369 CTCGCGAGC 378  
DB 915 GSGSGSGSVS 924

## RESULT 7

DR741095/c 1133 bp mRNA linear EST 18-JUN-2005  
LOCUS FGAS001026 Triticum aestivum FGAS: Library 2 Gate 3 Triticum  
DEFINITION aestivum cDNA, mRNA sequence.

ACCESSION DR741095  
VERSION DR741095.1 GI:70970528  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

## REFERENCE

1 (bases 1 to 1133)  
AUTHORS Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Latroche, A., Links, M.G., McCarthy, E.L., Monroy, A., Muzek, I., Nilsson, D., Peniket, C., Roach, J.L. and Sarhan, F.

TITLES Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
JOURNAL Unpublished (2003)  
CONTACT: Patrick Gulick  
Plant Molecular Biology  
Concordia University, Department of Biology  
7141 Sherbrooke St. West, Montreal, Quebec H4B 1K6, Canada  
Tel: 514 848 2424 Ext 3407  
Fax: 514 848 2881  
Email: [pgulick@alcor.concordia.ca](mailto:pgulick@alcor.concordia.ca)

This sequence is the direct result of the Base calling software phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [32,747].  
Plate: L280204 row: J column: 19.

## FEATURES

Source  
1..1133  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum FGAS: Library 2 Gate 3"  
/note="Organ: Crown and leaf, Vector: pCMV.SPORT6, Aerial"

## ORIGIN

QY	24	CCGAGCGCCGACACCTGTGCAAGATCTCCGCGCGGTAGCAGGTGTCCTTAGAGCGGCT	83
Db	586	CTCTATGCGCAATACCTGCTCGACAACTCTCGGCCCTTAGAGTTCCTGCAAAACCTCC	527
QY	84	GCGGCCGACGCGCAGAACTGTGCAACGCGAAGCGCACCCGCGCTTGAGACGAGCGAGGC	143
Db	526	CTTCCCAATGTGGACAGAAACGACAAACGCAAGGGGACGCCAACTGTGGAATGAACGAGGC	467
QY	144	GGTGTAGCC-----GTGTTGGCAGCGAATCCGCAACCTCTCGATGAATATGCC	191
Db	466	TGTATGTTTAGAGCGCTGCGCCCTGTCTGTCAACGGGGATATCCCAACGATCTGTGTT	407
QY	192	GTGCAGCAGCTCGAACAGAACTTTGTGCT	221
Db	406	GTCTCTCAGCTTGAGAGTATTTGTGTGT	377

## RESULT 8

LOCUS	BI955767/c
DEFINITION	717 bp mRNA linear EST 19-OCT-2001
ACCESSION	HYSMEM0024G19f Hordeum vulgare green seedling EST library HYCDNA00014 (Blumeria infected) Hordeum vulgare subsp. vulgare CDNA clone HYSMEM0024G19f, mRNA sequence. c1955767

ACCESSION	BI955767
VERSION	BI955767.1
KEYWORDS	GI:16302370 EST.

SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.

REFERENCE  
AUTHORS  
1 (bases 1 to 717)  
Wing, R., Close, T.J., Kleinholz, A., Wise, R., Chin, A., Begum, D.

Frishch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.  
 Simmons, T., Oates, D., and Main, D.

**TITLE**  
Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library

**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Wing RA

Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
T. J. Jordan, tjjordan@clemson.edu

Tel.: 864 656 7288  
 Fax: 864 656 4293  
 Email: [rwjng@clermson.edu](mailto:rwjng@clermson.edu)  
 Total hg bases = 310  
 Seq primer: **ATTATACCTCCTACTAATAGGC**  
 High quality sequence start: 4  
 High quality sequence stop: 551

## FEATURES

```

source
1. .717
/organism="Horddeum vulgare subsp. vulgare"
/mol type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMDm0024G191"

```

“tissue type=“green seedling leaf”  
/lab host=“T0C121”  
/clone.lib=“Hordeum vulgare green seedling EST library  
HVCDA0014 (Blumeria infected)”  
/note=“Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; Moxr (ma) plants were greenhouse grown in the R  
Wise lab at Iowa State University, Ames, IA; 7 day old  
green seedlings were infected with isolate 5874 of  
Blumeria graminis f. sp. hordei, and leaves were harvested  
24, 48 and 72 hr post-inoculation and snap frozen (Wise).  
In the T0 Close lab at the University of California,  
Riverside, total RNA was prepared from each sample pool,  
equal quantities of all three RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give pBluescript SK(-) cDNA phagemids  
(Chn). Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order  
this clone see <http://www.genome.clemson.edu/orders> Also  
see Close TJ, Wing R, Kleinofe A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(<http://wheat.ipw.usda.gov/gspages/bgn/31/cover.html>).”

Query Match	9.1%	Score	54.2	DB 3	Length	717
Best Local Similarity	58.5%	Pred. NO.	0.0044			
Matches 120; Conservative	0	Mismatches	73	Indels	12	Gaps

**Dy** 29 GGCGCAGCACTGTCGCACGATCTGCCGGTAGCAGGTGTCGCTTGAAGGCGCTCGGCGC 88  
||||| ||||| | |||||  
**Db** 557 GCGCCAACTGCTAGATTACTGTGAGCATTTAAGTTCCTTGCGCCCCCTCGTTCC 498  
||||| ||||| | |||||

QY 89 CCCACGCGAGAACTGTGCACGCGAAGCGCGCACCCCGCTTGAGACGAGCGAGCGGTGT 148  
 Db 497 CTGTGCGACAAACGACGACACGACGAGGCGCGCCCGCTGTGATGAAACGACGAGCGGTGA 438

Oy 149 AGCC-----GTCGTGACACGAGATCCGACCGTGTGATGATAAATGCGCTGCA 196  
 |||||  
 Db 437 GCTTAGGCGTGCCTCTGTGTTCACAGGAGATCCGACCGTGTTCACCAACCTGTGTCT 378

QY	197	CCAGCTGGAACAGGA	CTTGTGCT	221
Db	377	CCAGCTTGAGAA	GTATCTTGTGCT	353

## RESULT 9

	CY948578	655 bp	mRNA	linear	EST 25-JAN-2005
LOCUS	PtPrvP_11064	zoospores,	purified	Phytophthora infestans	cDNA, mRNA
DEFINITION	sequence.				

ACCESSION CV948578  
VERSION CV948578.1 GI:58138334  
KEYWORDS EST.

SOURCE ORGANISM	(potato late blight agent)
Phytophthora infestans	
Phytophthora infestans	

1 Eukaryota; stramenopila; Phaeophyceae; Phaeophytophthora.

## AUTHORS

Kelkar, H., Fong, A.M., Gates, K., Roberts, S., Yatzkan, E., Gaffney, T., Law, M., Teeta, A., Torto, A., Zhang, M., Zheng, L., Mueller, R., Windass, J., Binder, A., Birch, P.R.J., Gisi, U., Govers, F., Gow, N.A., Mauch, P., van West, P., Waugh, M.E., Yu, J., Bollner, T., Kamoun, S., Lam, S.T. and Juddelson, H.S.

**TITLE**  
Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi

**JOURNAL**  
Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)

**PUBMED**  
15782637

**COMMENT**  
Contact: Judelson HS  
Department of Plant Pathology  
University of California  
Webber Hall, Riverside, CA 92521, USA  
Tel: 909 787 4199  
Fax: 909 787 4294  
Email: howard.judelson@ucr.edu.

**FEATURES**  
source  
1..655  
/organism="Phytophthora infestans"  
/mol\_type="mRNA"  
/strain="88069"  
/db\_xref="taxon:4787"  
/sex="M"  
/clone\_lib="zoospores, purified"  
/note="Vector: pSPORT1"

**ORIGIN**  
Query Match 9.0%; Score 53.8; DB 8; Length 655;  
Best Local Similarity 55.7%; Pred. No. 0.0055;  
Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 37 ACTGCTGCACGATCTCCGCCGCTGACAGGTGTGCTTGAAGCCCTCCGCCGACGCC 96  
DB 491 AGCTGCTCCAGATCTCCGCACTGCAGAGTGTGCGGAGATACCAAGTGTGCGTGC 432  
QY 97 CAGAACTGGACCGGACGCGGACCGGCTGAGACGACGACGCGGATGACCGCTGCG 156  
DB 431 CAGAACTGGACCGGACCGGCTGAGACGCGGCTGAGACGCGGATGACCGCTGCG 372  
QY 157 TGGCAGCGATCCGACCGCTGCTGATGAATTCGCTGACCGCTGACGACGAACTTT 216  
DB 371 GTGTGGCGCATGATCAAGCGCTTGCATCTGTCGTCATCTTGAAGTTTAAGCAGTAATTC 312  
QY 217 GTGCT 221  
DB 311 ACCGT 307

**RESULT 10**  
CV948570 671 bp mRNA linear EST 25-JAN-2005  
LOCUS PVXpvb.12056 zoospores, purified *Phytophthora infestans* CDNA, mRNA  
DEFINITION  
ACCESSION CV948570 GI:58138326  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Phytophthora infestans (potato late blight agent)  
Phytophthora infestans  
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
Phytophthora.  
1 (bases 1 to 671)  
Randal, T., Dwyer, R.A., Huitema, E., Beyer, K., Cvitanich, C.,  
Kelkar, H., Fong, A.M., Gates, K., Roberts, S., Yalcin, E., Gaffney, T.,  
Law, M., Teale, A., Torto-Alalibo, A., Zhang, M., Zheng, L., Mueller, E.,  
Windass, J., Binder, A., Birch, P.R.J., Gisi, U., Govers, F., Gow, N.A.,  
Mauch, F., van Meer, P., Maugh, M.E., Yu, J., Bolter, T., Kamoun, S.,  
Lam, S.T., and Judelson, H.S.  
Large-scale gene discovery in the oomycete *Phytophthora infestans*  
reveals likely components of phytopathogenicity shared with true  
fungi

**JOURNAL**  
Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)

**PUBMED**  
15782637

**COMMENT**  
Contact: Judelson HS  
Department of Plant Pathology  
University of California  
Webber Hall, Riverside, CA 92521, USA  
Tel: 909 787 4199

**FEATURES**  
source  
1..671  
/organism="Phytophthora infestans"  
/mol\_type="mRNA"  
/strain="88069"  
/db\_xref="taxon:4787"  
/sex="M"  
/clone\_lib="zoospores, purified"  
/note="Vector: pSPORT1"

**ORIGIN**  
Query Match 9.0%; Score 53.8; DB 8; Length 671;  
Best Local Similarity 55.7%; Pred. No. 0.0055;  
Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 37 ACTGCTGCACGATCTCCGCCGCTGACAGGTGTGCTTGAAGCCCTCCGCCGACGCC 96  
DB 448 AGCTGCTCCAGATCTCCGCACTGCAGAGTGTGCGGAGATACCAAGTGTGCGTGC 389  
QY 97 CAGAACTGGACCGGACGCGGACCGGCTGAGACGACGACGCGGATGACCGCTGCG 156  
DB 388 CAGAACTGGACCGGACCGGCTGAGACGCGGCTGAGACGCGGATGACCGCTGCG 329  
QY 157 TGGCAGCGATCCGACCGCTGCTGATGAATTCGCTGACCGCTGACGACGAACTTT 216  
DB 328 GTGTGGCGCATGATCAAGCGCTTGCATCTGTCGTCATCTTGAAGTTTAAGCAGTAATTC 269  
QY 217 GTGCT 221  
DB 268 ACCGT 264

**RESULT 11**  
BF265573 849 bp mRNA linear EST 23-OCT-2001  
LOCUS HV\_CEA0012L01f Hordeum vulgare seedling green leaf EST library  
DEFINITION  
ACCESSION BF265573  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Hordeum vulgare subsp. vulgare  
EST.  
Hordeum vulgare subsp. vulgare  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
1 (bases 1 to 849)  
Wing, R., Close, T.J., Kleinhofe, A., Wise, R., Wei, F., Begum, D.,  
Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J.,  
Choi, D.W., Fenton, R.D., Oates, R., and Main, D.  
Development of a genetically and physically anchored EST resource  
for barley genomes: Blumeria infected incompatible (MLA13)  
seedling leaf cDNA library  
Unpublished (2001)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hg bases = 448  
Seq primer: AATTACCTCACTAAAGG  
High quality sequence stop: 578.  
Location/Qualifiers  
1..849  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="C16155 (MLA13)"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"

**JOURNAL**  
Mol. Plant-Microbe Interact. 18 (3), 229-243 (2005)

**PUBMED**  
15782637

**COMMENT**  
Contact: Judelson HS  
Department of Plant Pathology  
University of California  
Webber Hall, Riverside, CA 92521, USA  
Tel: 909 787 4199

/clone="HV\_CEA0012L01f"  
 /tissue type="seedling green leaf"  
 /lab host="TUC121"  
 /clone lib="Hordeum vulgare seedling green leaf EST  
 library HVCN0004 (Blumeria challenged)  
 /note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 C.I. 1615 (M1a13) plants were greenhouse grown in the R  
 Wise lab at Iowa State University, Ames, IA; 7 day old  
 green seedlings were challenged with isolate A27  
 (AvrM1a13) of Blumeria graminis f. sp. hordei, and leaves  
 were harvested 20 and 24 hr post-inoculation and snap  
 frozen; uninoculated leaves were harvested 20 hr  
 post-inoculation (Wei, Wise). In the TJ Close lab at the  
 University of California, Riverside, total RNA was  
 prepared from each sample pool, equal quantities of all  
 three RNA pools were combined, poly(A) RNA was purified  
 from the mixture, one cDNA library was made, and 1 million  
 phagemids (Choi, Close). Phagemids were plated and picked  
 at the Clemson University Genomics Institute (CUGI)  
 (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
 preparations, DNA sequencing and sequence analysis were  
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,  
 Oates, Rambo, Main). The sequence has been trimmed to  
 remove vector sequence and contains a minimum of 100 bases  
 of phred value 20 or above. For more details on library  
 preparation and sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order  
 this clone see <http://www.genome.clemson.edu/orders> Also  
 see Close TJ, Wing R, Kleinborts A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (<http://wheat.pw.usda.gov/g95pages/bgn/31/cover.html>).

## ORIGIN

Query Match 9.0%; Score 53.8; DB 2; Length 849;  
 Best Local Similarity 58.6%; Pred. No. 0.0056;  
 Matches 119; Conservative 0; Mismatches 72; Indels 12; Gaps 1;  
 QY 31 GCCAGACCTGCTGCACGATCTGCCGAGTAGAGTGTGCTTGAAGCCGTCGCGCC 90  
 DB 517 GCCAACCCTGCTGCAGTATCTGTGAGCTTTTAAAGTTCCTTGCAGAACCTCTCCCTCC 458  
 QY 91 GAGCCGACAACTGCGACGAGGAGGCGCACCCGCTTGAGACGACGCGGCGGTAG 150  
 DB 457 GTGGCACAAAAGAGGAGGAGGCGACCGACCTGTGATGAAACGCGGCGGTAGC 398  
 QY 151 CC-----GTCGTGAGGAGGAGGATCCGACCGTCTCGATGAATTGCGCGTAGC 198  
 DB 397 CTAGGGGTGCGCCCTGTCTGTGACAGGAGATCCCTACCGTCTTCACCAACCTGTGTCTCC 338  
 QY 199 AGCTCGAAGAGAACTTGTGCT 221  
 DB 337 AGCTGAGAAAGTATCTTGTGTGT 315

RESULT 12  
 CK152637 857 bp mRNA linear EST 05-DEC-2003  
 LOCUS FGAS035711 Triticum aestivum FGAS: Talt3 Triticum aestivum cDNA,  
 DEFINITION mRNA sequence.  
 ACCESSION CK152637  
 VERSION CK152637.1 GI:38971884  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Triticum.  
 1 (bases 1 to 857)  
 Allard, F., Crosby M.L., Danyluk, J., Sudes, F., Frick, M., Gaudet, D.,  
 Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,  
 Links, M.G., McCarthy, E.L., Monroy, A., Murak, I., Nilsson, D.,

TITLE Penniket, C., Roach, J.L. and Sathan, F.  
 JOURNAL Functional Genomes of Abiotic Stress in Wheat and Canola Crops  
 COMMENT Unpublished (2003)  
 CONTACT Contact: Wm L Crosby  
 BIOINFORMATICS Bioinformatics  
 INSTITUTION University of Saskatchewan, Department of Computer Science  
 LOCATION 1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
 SASKATCHEWAN, S7N 5A9, Canada  
 TEL: 306 966 1769  
 FAX: 306 966 2033  
 EMAIL: fgas\_est@cs.usask.ca

FEATURES  
 source  
 1..857  
 location/Qualifiers  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Wheat line CI 14106"  
 /db\_xref="taxon:4565"  
 /lab\_host="DH5 alpha"  
 /clone\_lib="Triticum aestivum FGAS: Talt3"  
 /note="Organ: Crown; Vector: pGEM-T; SSH (suppression  
 subtractive hybridization) cDNA library from genotype  
 CI4106 cold hardened at 2 C for 21 days and 49 days  
 (equal amount of cDNA pooled together before subtraction,  
 tester) and subtracted against genotype Norstar cold  
 hardened at 2 C for 1 day (24 H (driver). Nitro-pyrole  
 anchored oligo-dT priming and non-directional cloning."  
 Plate: Talt350 row: A column: 01.

## ORIGIN

Query Match 8.9%; Score 53.4; DB 7; Length 857;  
 Best Local Similarity 57.6%; Pred. No. 0.0071;  
 Matches 121; Conservative 0; Mismatches 77; Indels 12; Gaps 1;  
 QY 24 CCGAGCGCCAGACCTGCTGCACGATCTGCCGAGTAGAGTGTGCTTGAAGCCGTC 83  
 DB 452 CTCTATGCGCAAGACTGCTGCACGATCTGTGAGCTTAAAGTTCCTGCGAACCTCC 393  
 QY 84 GCGCGCCGACGCGCAAACTGCGACGCGGAGGCGCACCCGCTTGAGACGACGAGC 143  
 DB 392 CTTCGCAAGTGCGACAGAAAGAGGAGGCGGCGGCGGACCTGTGAGCAAGAACAGAGGC 333  
 QY 144 GGTGTAGCC-----GTCGTGAGGAGGAGGATCCGACCGTCTCGATGAATTGCG 191  
 DB 332 GGTAGCTTAAAGCGTGTGCTGTGACAGGAGATCCCGCTCTCCACCAACGTGT 273  
 QY 192 GTCGACCACTGGAACAGAACTTGTGCT 221  
 DB 272 GTCTCCCAACTTGAGAAAGTATCTTGTGTGT 243

RESULT 13  
 BE415712 339 bp mRNA linear EST 24-JUL-2000  
 LOCUS MML038.G04000426 ITCC MML Wheat Root Library Triticum aestivum cDNA  
 DEFINITION clone MML038.G04, mRNA sequence.  
 ACCESSION BE415712  
 VERSION BE415712.1 GI:9413558  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Triticum.  
 1 (bases 1 to 339)  
 Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T.,  
 Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A.,  
 Gustafson, P., Herrmann, R.G., Holton, T., Jaccuemin, J.M., Jia, J.,  
 Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.C., McGuire, P.,

REFERENCE  
 AUTHORS



TITLE  
Ogilhara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrelle, M., Warburton, M. and Wenzel, G.  
International Triticale EST Cooperative (ITREC): Production of Expressed Sequence Tags for Species of the Triticaceae  
Unpublished (2000)  
Contact: Warburton M  
Applied Biotechnology Center, CIMMYT  
Apdo. Postal 6-641, 06600 Mexico DF MEXICO  
Tel: 52-5-7269091 ext 1381  
Fax: 52-5-7267558/59  
Email: mwarburton@cgmnet.com  
International Triticale EST Cooperative (ITREC)  
http://wheat.pw.usda.gov/genome.  
Location/Qualifiers  
1..339  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Atlas"  
/db\_xref="taxon:4565"  
/clone="MM1038.G04"  
/tissue\_type="root"  
/dev\_stage="8 day old"  
/clone\_id="ITREC MML Wheat Root Library"  
/note="vector: pYES2 (Invitrogen); 0.5-1.5 Kbp average insert size."

ORIGIN  
Query Match 8.8%; Score 52.8; DB 2; Length 339;  
Best Local Similarity 58.1%; Pred. No. 0.0096;  
Matches 118; Conservative 0; Mismatches 73; Indels 12; Gaps 1;

Db 31 GCCAGCACTGCTGACGATCTGCGCCGCTAGCAGGTCTGCTGAGCGCCGCGGCC 90  
317 GGCANACCTGCTGACATCTGCGCCCTTAAAGTTCCTGCGAACCTCTCCCA 258  
Qy 91 GACGCGCAGAACTGCGACGCGACGCGCCGCTGAGCAGACGCGAGCTGAG 150  
Db 257 GTGCGACGAAACGACGACGCGCGGCGACGCTGTGACGAAACGACGCGTGGC 198  
Qy 151 CC-----GTGCTGGCGACGATCCGACCGCTCTCGATGAATTGCCGTGACC 198  
Db 197 CTAGGCGTGCCTGCTGCAACGGGAGATCCCACTGCTCCACCACTGTGTCTCC 138  
Qy 199 AGCTGGAACGAACTTGTGCT 221  
Db 137 AACTGAGAGATCTGTGTGT 115

RESULT 14  
LOCUS CD922535 528 bp mRNA linear EST 15-JUL-2003  
DEFINITION G750.103J05F010528 G750 Triticum aestivum cDNA clone G750103J05,  
mRNA sequence.  
ACCESSION CD922535  
VERSION CD922535.1 GI:32770299  
KEYWORDS EST  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticaceae; Triticum.  
1 (bases 1 to 528)  
Genoplane.  
Genoplane, a major partnership french program in plant genomics  
Unpublished (2003)  
Contact: Genoplane  
Genoplane  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genome programme 'Genoplane' (http://www.genoplane.com  
and http://genoplane-info.infobiogen.fr).

FEATURES  
source  
Location/Qualifiers  
1..528  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="rectal"  
/db\_xref="taxon:4565"  
/clone="G750103J05"  
/tissue\_type="grain (750 degrees per day after pollination)"  
/clone\_id="G750"

ORIGIN  
Query Match 8.8%; Score 52.8; DB 6; Length 528;  
Best Local Similarity 57.6%; Pred. No. 0.0098;  
Matches 121; Conservative 0; Mismatches 77; Indels 12; Gaps 1;

Qy 24 CCGAGCGCGACCACTGCTGACGATCTGCGCCGCTAGCAGGTGCTTGAAGCCGTC 83  
Db 270 CTTATATGCCAATCTGCTGACATCTGCGCCCTTGAAGTTCCTCGAACCTCTCC 211  
Qy 84 GCGGCGCGACGCGCAACTGCGACGCGAAGGCGCACCCGCTGAGACGACGAGC 143  
Db 210 CTTCCAGTGCACAGAACGACGACGCGAGCGACCACTTGCATGAACGCAAGT 151  
Qy 144 GGTGTAGCC-----GTGCTGGCGACGATCCGACCGCTCTGATGAATTGCC 191  
Db 150 TGTAGCTTAGAGCGCTGCGCCCTGTCTCAAGGGGATCCACCGTCTCCACAGTCTGT 91  
Qy 192 GTGACCACTGCGAACGAACTTGTGCT 221  
Db 90 GTCTCCAGCTTGAGAGATCTGTGTGT 61

RESULT 15  
LOCUS CK124607 771 bp mRNA linear EST 01-MAR-2004  
DEFINITION BE51824109M05 BE51824 Hordeum vulgare subsp. vulgare cDNA clone  
MPMG2010M055 5-PRIME, mRNA sequence.  
ACCESSION CK124607  
VERSION CK124607.1 GI:44807609  
KEYWORDS EST  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticaceae; Hordeum.  
1 (bases 1 to 771)  
Kramer, A., Fellner, T., Possling, A., Radchuk, V., Weschke, W.,  
Buerkle, L., and Kersten, B.  
Application of the protein microarray technology for the  
identification of expression library derived target proteins for  
barley protein kinase CK2  
Unpublished (2003)  
Contact: Birgit Kersten\* and Winfriede Weschke\*\*  
\*Plant Protein Chip Group, Department Lehrich, \*\*Department  
Molecular Genetics, Gene Expression Group  
\*Max-Planck-Institute for Molecular Genetics, \*\*Institute of Plant  
Genetics and Crop Plant Research Gatersleben  
\*Innestr. 73, D-14195 Berlin, Germany, \*\*Corrensstrasse 3, D-06466  
Gatersleben, Germany  
Tel: \*\*49(0)30/84131648, \*\*49(0)394825500  
Fax: \*\*49(0)30/84131128, \*\*49(0)394825237  
Email: \*kersten@molgen.mpg.de, \*\*weschke@ipk-gatersleben.de  
Insert Length: 771 Std Error: 0.00  
Plate: 9 row: M column: 5  
Seq primer: PDE65.  
Location/Qualifiers  
1..771  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Barke"  
/db\_xref="taxon:4565"  
/sub\_species="vulgare"  
/db\_xref="GABI:945637"



```
/db_xref="taxon:112509"
/clone="MPMG2010M059"
/issue_type="embryosac"
/dev_stage="0-10 DAF (days after flowering)"
/lab_host="E. coli, SCS-1/PSK11"
/clone_1lb="BES1824"
/notes="Vector: pOE3ONST (AF074376); Site 1: Sali; Site 2:
NotI; 0-10 DAF (days after flowering), cDNA synthesis
using Bluescript II XR cDNA-library construction kit
(Stratagen) with an oligo(dT)-primer containing NotI
restriction site and a Sali adapter (Invitrogen). The main
library of 21500 clones was rearranged into the sublibrary
BES 1824 containing 4100 putative expression clones. Note:
Due to a cloning artefact caused by the kit, in most cases
the Sali site is NOT present, as well as the Sali Adapter
used for cloning. To excise the insert, restriction sites
upstream Sali should be used (e.g. BamHI). Average insert
size is 1 kb. Library generation and sequencing was
granted in context of GABI; data are also accessible at
https://gabi.rzpd.de"
```

## ORIGIN

```
Query Match      8.8%; Score 52.8; DB 7; Length 771;
Best Local Similarity 57.6%; Pred. No. 0.01;
Matches 121; Conservative 0; Mismatches 77; Indels 12; Gaps 1;

QY 24 CCCGAGGCGCAGCACTCTGCAAGATCTGCCCGGTAGCAAGTGTCTTGAAGCCGTC 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 413 CTCTATAGGCCAACAACCTGCTGACTATCTGTGCCCCCTTAAGTTCTTACCAACCTCC 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 84 GCGGCGCGAGCGAGACTGCGACGCGAGGCGCACCCCGCTGAGACGAGACGACGCG 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 353 CTTCCTGTGCGCAAAAACAGACGACGAGGCGACCGACCTGTATGAAACGACAGGC 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 GGTGTAGCC-----GTCTGTGCGACGATCCGACCGCTCTGATGAATATGCC 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 293 GGTAGGCTTAGGCGTGCCTGTCTGTGACGAGGAGATCCATACGCTCTCCACCAACCTGT 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 192 GTGACCAAGCTGAAACGAACTTTGTGCT 221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 233 GTCTTCAGCTTAGAAGTATCTTGTGTGT 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: April 7, 2006, 14:04:07  
Job time : 2303.02 secs

**This Page Blank (uspto)**

QY	26	CGAGCCGACGACCTGCTGCAGCATCTCGGCCGCGTAGCGAGTGTGCTTGAAGCCGTCGC	85
Db	3184316	CGCCGGACGACACTGCTTCAGAGATCTCGCCCTCGATGAGTTGCGAGTCAATCGCCCT	3184375
QY	86	GCGCCGACGCGCGCAAACTGCGACGCGAAGCGCACCCGCGCTGAGACGAGACGCGCG	145
Db	3184376	GCGCAGTGTGCGACGAAACGGCGAAGCCATGTGCGCAGCGCGCTCGAGGAATATGACACG	3184435
QY	146	TGTAGCCGTGTGCGCGACGCGATCCGACCGTCTCGATGAATATGCGGTGCAACGACTCGA	205
Db	3184436	TGTTGCGCCCGCGATAGCGGATACGACACCGATTGGAACATGTATACGTTCACGCGCCGCC	3184495
QY	206	ACAGAACTTTGTGTGTGCG	226
Db	3184496	ACACGCTTTTCAGATGTGCG	3184516

## RESULT 2

US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103, 840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 13.0%; Score 77.8; DB 3; Length 4411529;  
Best Local Similarity 61.7%; Pred. No. 2e-09;  
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 26 CGAGCCGACGACTGCTGCAAGATCTGCGCCGTAAGAGTGTGCTTGAAGCCCTGCGC 85  
DB 3190207 CGCCGCGACGCGACTGCTGCGAGATCTCCGCTGATGAGTTCGATCCTCGCCCT 3190266  
QY 86 GCGCCGACGCGGAGACTGCGACGCGGAGGCGACCCCGCTGAGACGACGAGCGCG 145  
DB 3190267 GCGCAGTGGCACAGACGCGGAGGAGCCATGCGACCGCGCTTGGGAGAAATGCAAGCG 3190326  
QY 146 TGTAGCCCTGCTGCGAGATCCGACCCGCTCTCGATGAATGCGCGTGAACCACTGCA 205  
DB 3190327 TGTTCGCGCGGAGATGAGCATGACGACGATTCGAAATGATGATGACGCGCGCCG 3190386  
QY 206 ACAGGAACCTTGTGCTGCTGCGC 226  
DB 3190387 ACAACGCTTTCGAGTCTGCGC 3190407

## RESULT 3

US-09-902-540-9688/c  
Sequence 9688, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 9688  
LENGTH: 1080  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-9688

Query Match 13.0%; Score 77.6; DB 3; Length 1080;  
Best Local Similarity 61.3%; Pred. No. 4.2e-10;  
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 34 AGCACTGCTGACGATCTGCGCCGATGACAGTGTGCTTGAAGCGCGTGGCGCCGAC 93  
DB 431 AGTACTGCGCCACGATTTGCGCGGCGGTGAGTGGCGGTGAGAGCCCATCTGCGCGGTG 372

QY 94 GCGGAGAACTGCGACGCGGAGGCGCAACCCCGCTGAGACGAGCGAGCGGTAGCCG 153  
DB 371 GCGGAGAAAGCGAGCGCCATGCGCGACGCGCGCTGCTGTGATGACACCGTGGCGCGG 312  
QY 154 TCGTGGGACGAGATCCGACCGTCTCGATGAATGCGCGTGAACGAGCTGAACGAGAC 213  
DB 311 CCTTGAACCGCATGACGACGATTGCAATGCTGCGCGTGTCAAGTCCAGACGACGACG 252

QY 214 TTTGTGCTGCGCTTCCCGTGGTG 237  
DB 251 TTGTGCGTGAAGCGGTGCTGCTG 228

## RESULT 4

US-09-902-540-1099/c  
Sequence 1099, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 1099  
LENGTH: 13751  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-1099

Query Match 13.0%; Score 77.6; DB 3; Length 13751;  
Best Local Similarity 61.3%; Pred. No. 7e-10;  
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 34 AGCACTGCTGACGATCTGCGCCGATGACAGTGTGCTTGAAGCGCGCTGCGCCGAC 93  
DB 5590 AGTACTGCGCCACGATTTGCGCGGCGGTGAGTGGCGGTGAGAACCCCATCTGCGCGGTG 5531  
QY 94 GCGGAGAACTGCGACGCGGAGGCGCACCCCGCTGAGACGAGCGAGCGGTAGCCG 153  
DB 5530 GCGGAGAAAGCGAGCGCCATGCGCGACGCGCGCTGCTGTGATGACACCGTGGCGCGG 5471  
QY 154 TCGTGGGACGAGATCCGACCGTCTCGATGAATGCGCGTGAACGAGCTGAACGAGAAC 213  
DB 5470 CCTTGAACCGCATGACGACGATTGCAATGCTGCGCGTGTCAAGTCCAGACGACGACG 5411  
QY 214 TTTGTGCTGCGCTTCCCGTGGTG 237  
DB 5410 TTGTGCGTGAAGCGGTGCTGCTG 5387

## RESULT 5

US-09-902-540-3050/c  
Sequence 3050, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825

```
; SEQ ID NO 3050
;
; LENGTH: 792
;
; TYPE: DNA
;
; ORGANISM: Myxococcus xanthus
US-09-902-540-3050
```

Query Match	12.6%;	Score 75;	DB 3;	Length 792;
Best Local Similarity	58.9%;	Pred. No. 1.8e-09;		
Matches 129;	Conservative	0;	Mismatches 90;	Indels 0;
			Gaps	0;

Oy 3 CACGAGCCCTGTGAGCGCCGGGCCCCGAGCGCCAGCACTCTGTGACGATCTCCGCCGCTAG 62  
 Db 459 CTCGTTCTTGGCAACCTCGCGGTTACCGCGTGAACCTTGGCCACAATCTCGTGGGGGT 400  
 Oy 63 CAGGTGTGCTTGAAGCCGCTCGCGGCCGACGCGCAGAACTTGACGCGAAGCGCACCC 122  
 Db 399 CAGGTTCGCTTGAAGCCCCATGATGTGCCCCGTCAATGACGAAAGCCGCAAGCCATGCGGAGCC 340  
 Oy 123 CGCCTGAACAGAAAGCGAGGCGGTGTGAGCCGCTGTGCGACAGGATCCGACCGCTTCGAT 182  
 Db 339 CACCTCGCTGGAACAAGCAACAGTCCTCGGCGTCTCTCGGCGCATGTAGACGATTCGAT 280  
 Oy 183 GAAATTGCGGTGACCAAGCTGAAACAGAACTTTGTCGT 221  
 Db 279 GTAGGCGCCGTTCCGCGTCTTCCACCGGACTTGATAGT 241

RESULT 6  
US-09-902-540-793  
; Sequence 793, Application US/09902540

; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10 (15849) B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 793  
 ; LENGTH: 6713  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 ; JS-09-902-540-793

Query Match	12.6%	Score 75;	DB 3;	Length 6713;
Best Local Similarity	58.9%	Pred. No. 2.7e-09;		
Matches 129;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0

QY CAGGACCTGTGAGCGCGGGGCGGAGGCGCAGACCTGTGCGACGATCTGCGCCGGTAG 62  
Db CTGCTTCTTGCGAAGCTCGGGTTCACGCGGTGACCTGGGCGACAAATCTCGTGGGGGT 395  
QY CAGGTGTGCTTGAAGGCGGTGCGGCGCGACGCGAGAACTGGGACGCGAGCGCACCC 122  
Db CAGGTGTGCTTGAAGGCGGTGCGGCGCGACGCGAGAACTGGGACGCGAGCGCACGCC 455  
QY CGGCTTGAGACGAGACGCGAGGCGGCTGTAGCCGTGCTGCGACGATTCGCAACCGTCTGAT 182  
Db CACTCTCGGTGACACGCAACAAGTCTGTGGGTCTCGGTGGGCAATGTAAGACGATTCGAT 515  
QY GAATTTGCGGTGACCAAGCTTCGACAGAGAACTTTGTCT 221  
Db GTAGCGCCGCTCCCGGCTTTCACCGGTACTTGAATGT 554

RESULT 7  
US-09-252-991A-6211  
; Sequence 6211, Application US/09252991A

```

: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OP INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OP INVENTION: AERGININOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: PRIOR FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 6211
: LENGTH: 456
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-6211

```

Query Match	10.9%;	Score 64.8;	DB 3;	length 456;
Best Local Similarity	60.0%;	Pred. No. 5.6e-07;		
Matches 108; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0

QY 36 CACCTGCTCAAGATCTCGCCGAGTAGAGAGGTTCGCTTGAAGCCGTGCGCGCCCAAGC 95

Db 144 CACCTGCGCGATTCACCTCGGCGGCGGTGAGGTCGCTGTGAAACCTGCTTGGCGGTGGA 203

QY 96 GCAGAACTTGGCAGCGGAGAGCGACCCCGCTGAGAGAGAGCGAGCGGTGAGCCGTG 155

Db 204 GCAGAACTTGCATTCAGGGCGCAGCCCGCTTGGAGAGACAGCAGAGGTGCGCGCTCC 263

QY 156 GTGCGCAGCGGATCCGCAACCGTCTCGATGAATTTGCGCTCGACCACTTCGACAGGAACTT 215

Db 264 GCCCTGAGGAGATGTCACAGGTCTCGAGCGCACTTCGCGGACGCCACCGGACCACTTCACTT 323

```

RESULT 8
US-09-252-991A-6337/C
: Sequence 6337, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 6337
: LENGTH: 855
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-6337

```

Query Match	10.9%;	Score 64.8;	DB 3;	length 855;
Best Local Similarity	60.0%;	Pred. No. 6.4e-07;		
Matches 108; Conservative	0;	Mismatches 72;	Indels 0;	Gaps 0

QY	3	CACCTGCTGCA	GAATCT	CGCCCGG	TAGAGGT	GTGCTGTTA	AGGCCGT	CGCGGCC	CGAAC	CG	95		
Db	436	CACCTGCGCGA	ATCACTT	CGCGCGG	CGGTGA	GGTCTGCTTT	GAATACCT	CGCTT	CGCGAT	GTGA	377		
QY	96	GCAGAACTG	CACCG	GAAGG	CGCACCC	CGCTTGA	AGACA	CGCAG	CGGT	GTAGCC	CTTC	155	
Db	376	GCAGAACTG	CACAT	CAAGGG	CGCAGCC	CGCTTGA	AGACA	CGCAG	CGGTG	CGCGT	CTTC	317	
QY	156	GTGGCGA	CGGAT	TCGCAC	CGCTCT	TGATGAA	ATTG	CGTCGA	CCAG	CTCGA	CAGGA	ATT	215
Db	316	GCCTGTGGG	GGATG	TACAG	CGGTCT	TGACG	CAATC	CGCGA	CGCAC	CGCGAC	CACTT	257	

## RESULT 9

US-09-252-991A-6419/c  
; Sequence 6419, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6419  
; LENGTH: 1164  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6419

Query Match 10.9%; Score 64.8; DB 3; Length 1164;  
Best Local Similarity 60.0%; Pred. No. 6.8e-07;  
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 36 CACCTGCTGCACGATCTCGCCGGTAGACAGTGTCCCTTGAGCCGCTCGCCGCCGACGC 95  
DB 456 CACCTGGCCGATCACTCGCGCGCGGTAGAGTCCCTGTTGAACCTGCTTCCGGTGA 397  
QY 96 GCAGAACTGGCAGCCGAGGCGCAACCCGCTGAGACGAGCAGGCGGTGACCGTC 155  
DB 396 GCAGAACTGCAATCCAGGCGCGACCCGCTGGGAGACACGACGAGGGTCCGCTGC 337  
QY 156 GTGGCGACGATCCGACCGCTCTGATGAATTTGCCGTGCACGCTGGAACGAACTT 215  
DB 336 GCCCTGGGGATGTACACGCTCTGACGCACTGCCGACCCGACCCGACCACTT 277

## RESULT 10

US-09-252-991A-6133  
; Sequence 6133, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6133  
; LENGTH: 1764  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6133

Query Match 10.9%; Score 64.8; DB 3; Length 1764;  
Best Local Similarity 60.0%; Pred. No. 7.3e-07;  
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 36 CACCTGCTGCACGATCTCGCCGGTAGACAGTGTCCCTTGAGCCGCTCGCCGCCGACGC 95  
DB 1525 CACCTGGCGGATCACTCGCGCGGTAGAGTCCCTGTTGAACCTGCTTCCGGTGA 1584  
QY 96 GCAGAACTGGCAGCCGAGGCGCAACCCGCTGAGACGACGACGAGGGGTGACCGCTC 155

DB 1585 GCAGAACTGCAATCCAGGCGCGACCCGCTGGAGAGACGACGAGGTCGCGCTCC 1644  
QY 156 GTGGCGACGATCCGACCGCTCTGATGAATTTGCCCTGACGACGATCGAAGAACTT 215  
DB 1645 GCCCTGGGGATGTACACGCTCTGACGCACTGCGGAGCCGACCCGACCACTT 1704

## RESULT 11

US-09-605-703B-2129/c  
; Sequence 2129, Application US/09605703B  
; Patent No. 6962989  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Krogger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zeider, Oskar  
; APPLICANT: Haberhauser, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL  
; FILE REFERENCE: BGI-129CP  
; CURRENT APPLICATION NUMBER: US/09/605,703B  
; PRIOR FILING DATE: 2000-06-27  
; PRIOR APPLICATION NUMBER: 60/142,764  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: 60/152,318  
; PRIOR FILING DATE: 1999-09-03  
; NUMBER OF SEQ ID NOS: 2934  
; SEQ ID NO 2129  
; LENGTH: 1221  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)..(1198)  
; OTHER INFORMATION: RXA01186  
US-09-605-703B-2129

Query Match 10.5%; Score 62.6; DB 4; Length 1221;  
Best Local Similarity 59.1%; Pred. No. 2.4e-06;  
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCACCTGCTGCACGATCTCGCCGGTAGACAGTGTCCCTTGAGGCGCTCGCGGCCGACG 94  
DB 566 GAACCTATCCACATCTCAACCGATTGAAGGTATACGATCCAAACCCCTGACCAATTG 507  
QY 95 CGCAGAACTGGCAGCGCAAGCGGACCCGCTGAGACGAGACGACGCGGTGAGCGGT 154  
DB 506 CACGAAATGGGACCGCATGCTCCGACCAAGCTTGGAGAAATTAACAAGCTGAGCGAT 447  
QY 155 CGTGGCAGCGATCCGACCGCTCTGATGAATTTGCCGTGACCAAGCTCGAACGAACT 214  
DB 446 CTGAATAGCGCATGAGAACAGACTCAAGTGCATATGAGAGCTTCATTAACGCTT 387  
QY 215 T 215  
DB 386 T 386

## RESULT 12

US-09-605-703B-2131/c  
; Sequence 2131, Application US/09605703B  
; Patent No. 6962989  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Krogger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zeider, Oskar  
; APPLICANT: Haberhauser, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL  
; FILE REFERENCE: BGI-129CP  
; CURRENT APPLICATION NUMBER: US/09/605,703B  
; PRIOR FILING DATE: 2000-06-27

;; PRIOR APPLICATION NUMBER: 60/142,764  
;; PRIOR FILING DATE: 1999-07-08  
;; PRIOR APPLICATION NUMBER: 60/152,318  
;; PRIOR FILING DATE: 1999-09-03  
;; NUMBER OF SEQ ID NOS: 2934  
;; SEQ ID NO 2131  
;; LENGTH: 1221  
;; TYPE: DNA  
;; ORGANISM: Corynebacterium glutamicum  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (101)..(1198)  
;; OTHER INFORMATION: RXA01186  
US-09-605-703B-2131

Query Match 10.5%; Score 62.6; DB 4; Length 1221;  
Best Local Similarity 59.1%; Pred. No. 2,4e-06;  
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCACCTGTCACGATCTCGCCGCGTATGACGATGTGTGAGGCGCGCGCGAG 94  
DB 566 GAACCTGATCCAGATCTCACCATTGAAGTTACGGTCCAAACCGCCTTGACAGTTG 507  
QY 95 CCGAAGACTGGACGCGGAAGCGCACCCGCTGAGACGACGCGCGGTGAGCCGT 154  
DB 506 CACAGATGGGACGCGCATGCGCGAACGCGCTGCGAAGAAATACACAGGTGAGAGAT 447  
QY 155 CGGCGGAGGATCCGCGCGCTTCGATGAATTCGCGTGCACGAGCTGACAGAACT 214  
DB 446 CTGAATATGGCATGAGAACTCAAGCAAGTGCATCATGAGCTTCATATAGCTCT 387  
QY 215 T 215  
DB 386 T 386

RESULT 13  
US-09-902-540-1116/C  
; Sequence 1116, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1116  
; LENGTH: 15377  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-1116

Query Match 10.3%; Score 61.4; DB 3; Length 15377;  
Best Local Similarity 48.6%; Pred. No. 8e-06;  
Matches 198; Conservative 0; Mismatches 206; Indels 3; Gaps 1;

QY 3 CACGAGCTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62  
DB 11342 CACCGCGCGGTCCGCTTCTCCCGGACCTGAGACCTGTCATATATCTCCAGGCTG 11283  
QY 63 CAGGTGCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122  
DB 11282 GAGGTTCCTTTGAAAGCCCACTTCCCGCTCATGCAAGAGTGCAGCGCGCGAGCC 11223  
QY 123 CGCTTGAAGACGACGACGCGGTGA---GCCGTGTGGCGACGAGTCCGACCGTCTC 179  
DB 11222 CACCTGGCTGAGAGCGCAATGACGTACTTTCATGGAAGATGGGATCGGACCGCTTC 11163

QY 180 GATGAATTCGCGTCGACGACCTGCAACGAACTTGTCTGTGCTTCCCTGTGCG 239  
DB 11162 GATACGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11103  
QY 240 ACTGCCCTCG 299  
DB 11102 GCGCTGACGATTTGCCAGCTTCGCGCATCTCCGCTGGGACGAGGTGTCTCCACGCG 11043  
QY 300 GCGCGATCTTTGCGCTGCGGATTAAGCTTGTGAAAGATGTGCGCTTCCGTTGCGC 359  
DB 11042 GCGCGGACCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10983  
QY 360 GCGGACCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 406  
DB 10982 GCGCGGACCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10936

RESULT 14  
US-09-902-540-2548/C  
; Sequence 2548, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 2548  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-2548

Query Match 10.2%; Score 60.8; DB 3; Length 966;  
Best Local Similarity 52.3%; Pred. No. 6.5e-06;  
Matches 134; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 3 CACGAGCTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62  
DB 342 CACCGCGCGGTCCGCTTCTCCCGGACCTGAGACCTGTCATATATCTCCAGGCTG 283  
QY 63 CAGGTGCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122  
DB 282 GAGGTTCCTTTGAAAGCCCACTTCCCGCTCATGCAAGATGCGACGCGCGCACAGCC 223  
QY 123 CGCTTGAAGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
DB 222 CACCTGCGTGAAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 163  
QY 183 GAAATTCGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 242  
DB 162 GATACGTGCG 103  
QY 243 GCGCTGCGCGAGGCTC 258  
DB 102 GCGCTGACGATTC 87

RESULT 15  
US-09-489-039A-4396/C  
; Sequence 4396, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS



FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 4396  
LENGTH: 1194  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-4396

Query Match 9.3%; Score 55.4; DB 3; Length 1194;  
Best Local Similarity 59.0%; Pred. No. 0.00015;  
Matches 95; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY	37	ACCTGCTGCACGATCTCGCCCGGTAGCAGGTGTGCTTGAGGCCGTGCGGCGCCGACGCG	96
DB	494	ACCTGCCGATTAATTTCGATACCCGACGTTGCGTTAAAGCCCTGCTGCGCGGTGGAG	435
QY	97	CAGAACTGGCAGCGGAGGCGGACCCCGCTGAGACGAGACGCGCGGTGTAGCCCTCG	156
DB	434	CAGATTGCACTCCAGCGCGGAGCTTACTTGGAGAGACGACAGGAGTGGCGGCT	375
QY	157	TGGGAGGAGTCCGACCGCTCTGATGAATTGCGGTGAC	197
DB	374	TCTTCCGGATTAAGACGCTCTCGACGCGCTGATGCCAAC	334

Search completed: April 7, 2006, 02:59:01  
Job time : 102.47 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:50 ; Search time 414.403 Seconds  
(without alignments)  
11913.100 Million cell updates/sec

Title: US-10-611-442-2\_COPY\_17000\_17596

Perfect score: 597  
Sequence: 1 aacacgagcctgtgacgcc.....cacgaacgcgcgcgcgcgc 597

Scoring table: IDENTITY\_NIC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 413468905 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: Published Applications NA Main:\*
- 2: /cgn2\_6/prodata/1/pubpna/us07\_PUBCOMB.seq:\*
- 3: /cgn2\_6/prodata/1/pubpna/us08\_PUBCOMB.seq:\*
- 4: /cgn2\_6/prodata/1/pubpna/us09\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/1/pubpna/us10\_PUBCOMB.seq:\*
- 6: /cgn2\_6/prodata/1/pubpna/us10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/1/pubpna/us10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/prodata/1/pubpna/us10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/1/pubpna/us10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/1/pubpna/us11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	17596	US-10-611-442-2	Sequence 2, Appli
2	77.2	12.9	1104	US-10-156-761-2608	Sequence 2608, Ap
3	77.2	12.9	9025608	US-10-156-761-1	Sequence 1, Appli
4	73.8	12.4	2256646	US-10-470-565-1	Sequence 1, Appli
5	62.6	10.5	1098	US-09-738-626-2214	Sequence 2214, Ap
6	62.6	10.5	3309400	US-09-738-626-1	Sequence 1, Appli
7	59.4	9.9	672	US-09-974-300-6756	Sequence 6756, Ap
8	56.4	9.4	2493	US-10-450-763-18200	Sequence 18200, A
9	56.4	9.4	2493	US-10-450-763-25444	Sequence 25444, A
10	55.6	9.3	594	US-09-974-300-2461	Sequence 2461, Ap
11	53.8	9.0	7685	US-10-194-163-1092	Sequence 1092, Ap
12	50	8.4	2715	US-10-156-761-2245	Sequence 2245, Ap
13	50	8.4	60872	US-10-915-740A-17	Sequence 17, Appli
14	50	8.4	2242716	US-10-915-740A-1068	Sequence 1068, Ap
15	48.8	8.2	1083	US-10-472-928-1455	Sequence 1455, Ap
16	48.8	8.2	1095	US-10-617-320-2203	Sequence 2203, Ap
17	48.8	8.2	5558	US-08-961-527-103	Sequence 103, App
18	48.8	8.2	5558	US-10-158-844-103	Sequence 103, App
19	48.8	8.2	2162598	US-10-472-928-4979	Sequence 4979, Ap
20	48.4	8.1	1339	US-10-425-114-35313	Sequence 35313, A
21	48.4	8.1	1460	US-10-425-114-5256	Sequence 5256, A
22	48.4	8.1	1832	US-10-425-114-2852	Sequence 2852, Ap
23	48.4	8.1	1832	US-10-425-114-20947	Sequence 20947, A

C 24	48.4	8.1	2170	US-10-425-115-161538	Sequence 161538,
C 25	48.4	8.1	2388	US-10-437-963-44594	Sequence 44594, A
C 26	48.2	8.1	573	US-10-487-901-2105	Sequence 2105, Ap
C 27	48.2	8.1	666	US-10-487-901-316	Sequence 316, App
C 28	48.2	8.1	1521	US-10-437-963-41296	Sequence 41296, A
C 29	47.4	7.9	400	US-08-781-986A-1342	Sequence 1342, Ap
C 30	47.4	7.9	400	US-10-329-624-1342	Sequence 1342, Ap
C 31	47.4	7.9	1513	US-10-437-963-80449	Sequence 80449, A
C 32	47.2	7.9	138203	US-10-819-386A-1	Sequence 1, Appli
C 33	46.8	7.8	900	US-10-437-963-84904	Sequence 84904, A
C 34	45.8	7.7	8955	US-10-857-625-51	Sequence 51, Appli
C 35	45.6	7.6	1659	US-10-437-963-93881	Sequence 93881, A
C 36	45.4	7.6	1792	US-10-425-114-15465	Sequence 15465, A
C 37	45.2	7.6	489	US-10-437-963-36501	Sequence 36501, A
C 38	45.2	7.6	1058	US-10-425-115-142879	Sequence 142879, A
C 39	45.2	7.6	1589	US-10-425-115-176166	Sequence 176166, A
C 40	44.8	7.5	1469	US-10-437-963-40940	Sequence 40940, A
C 41	44.8	7.5	2068	US-10-425-115-88284	Sequence 88284, A
C 42	44.8	7.5	3458	US-10-425-115-173488	Sequence 173488, A
C 43	44.6	7.5	948	US-10-437-963-27652	Sequence 27652, A
C 44	44.6	7.5	1413	US-10-156-761-2633	Sequence 2633, Ap
C 45	44.2	7.4	1542	US-10-156-761-4325	Sequence 4325, Ap

ALIGNMENTS

RESULT 1  
US-10-611-442-2  
; Sequence 2, Application US/10611442  
; Publication No. US20040203015A1  
; GENERAL INFORMATION:  
; APPLICANT: Hutchinson, C. Richard  
; APPLICANT: Katz, Leonard  
; APPLICANT: Reid, Ralph  
; APPLICANT: Hu, Zhihao  
; APPLICANT: Gramajo, Hugo  
; TITLE OF INVENTION: RECOMBINANT GENES FOR POLYKETIDE  
; TITLE OF INVENTION: MODIFYING ENZYMES  
; FILE REFERENCE: 306622009100  
; CURRENT APPLICATION NUMBER: US/10/611,442  
; CURRENT FILING DATE: 2003-06-30  
; PRIOR APPLICATION NUMBER: US 60/393,016  
; PRIOR FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 17596  
; TYPE: DNA  
; ORGANISM: Micromonospora Megalomicra  
; FEATURES:  
; NAME/KEY: misc.feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Genomic DNA  
US-10-611-442-2  
Query Match 100.0%; Score 597; DB 8; Length 17596;  
Best Local Similarity 100.0%; Pred. No. 2.6e-167;  
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACAGAGCCTGTGACGCGCGGCGGACGCGGACGAGCACTGCTGCAAGATCTGCCCGGT 60  
DB 17000 AACAGAGCCTGTGACGCGGCGGCGGACGCGGACGAGCACTGCTGCAAGATCTGCCCGGT 17059  
QY 61 AACAGAGCCTGTGACGCGGCGGCGGCGGCGGACGCGGACGAGCACTGCTGCAAGATCTGCCCGGT 120  
DB 17060 AACAGAGCCTGTGACGCGGCGGCGGCGGCGGACGCGGACGAGCACTGCTGCAAGATCTGCCCGGT 17119  
QY 121 CCGCGCTGAGAGGAGAGCGAGCGCGGTGACCGCTGCTGCGAGCGATCCGACCGTCTCG 180  
DB 17120 CCGCGCTGAGAGGAGAGCGAGCGCGGTGACCGCTGCTGCGAGCGATCCGACCGTCTCG 17179  
QY 181 ATGAATTCGCGTCGACGACGCTGCAACAGAACTTTGTGTGCTGCTTCCCTCGTGGCA 240



LENGTH: 2256646  
TYPE: DNA  
ORGANISM: Bifidobacterium longum  
US-10-470-565-1

Query Match 12.4%; Score 73.8; DB 7; Length 2256646;  
Best Local Similarity 60.0%; Pred. No. 2e-11;

Matches 123; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 26 CGAGCCGACGACCTGTCGACGATCTCCGCCGTGACGATGTCCTTGAAGCCCGCCG 85  
DB 1885220 CGGCCACACGACCTGTCGATTAATTTACCGGTGACATGTTACGGGTACGCCAAGCT 1885161  
QY 86 GGGCCGACGCGGAACTGCGACGCGAAGGCGACCCCGCTGAGACGACGAGCGG 145  
DB 1885160 TGCCTGCGGCGACGAGACGGGCACTTCATGCCGACGCCACTCTCGAGAGATGACAGCG 1885101  
QY 146 TGTAGCCGTCGTGGCGACGATCCGACCGCTCTGATGAAATTCCTCGACGACCTCGA 205  
DB 1885100 TGGTGGGGGTGCGGGTATCGCATTAAGCACCGATTGATGAGGAGCGCGTCGAAACGCTTC 1885041  
QY 206 ACAGGAACCTTGTCTGCTGCTTCC 230  
DB 1885040 ACAGGCTTGTGATGATGCTTCC 1885016

## RESULT 5

US-09-738-626-2214/c  
Sequence 2214, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 2214  
LENGTH: 1098  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-2214

Query Match 10.5%; Score 62.6; DB 3; Length 1098;  
Best Local Similarity 59.1%; Pred. No. 2.1e-08;

Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCAGCTGTCGACGATCTCCGCCGTGACGAGTGTGCTTGAAGCCGTCGCGCCGACG 94  
DB 466 GAACCTGATCCAGATCTCACCGATTGAAGGTTACGGTCCAAACGCCCTGACCAAGTTG 407  
QY 95 CGAGAACTGGACGCGGAGGCGCACCCGCTGAGACGACGCGGCTTACCGCT 154  
DB 406 CACAGATGGGACGCGCATGCGGCAACGCTGCGAAGAAATACACAGCGTGAACGAT 347  
QY 155 CGTGGGACGATCCGACCGTCTCGATGAATTCGCTGACCACTCGAAGCAAGACT 214

DB 346 CTGAATAGCGCATGAGAACAGACTCAAGAAAGTCCATCATGAGCTTCATTAAGCTCT 287  
QY 215 T 215  
DB 286 T 286

## RESULT 6

US-09-738-626-1  
Sequence 1, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 1  
LENGTH: 3309400  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 10.5%; Score 62.6; DB 3; Length 3309400;  
Best Local Similarity 59.1%; Pred. No. 4.4e-08;  
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 35 GCAGCTGTCGACGATCTCCGCCGTGACGAGTGTGCTTGAAGCCGTCGCGCCGACG 94  
DB 2135086 GAACCTGATCCAGATCTCACCGATTGAAGGTTACGGTCCAAACGCCCTGACCAAGTTG 2135145  
QY 95 CGAGAACTGGACGCGGAGGCGCACCCGCTGAGACGACGCGGCTTACCGCT 154  
DB 2135146 CACAGATGGGACGCGCATGCGGCAACGCTGCGAAGAAATACACAGCGTGAACGAT 2135205  
QY 155 CGTGGGACGATCCGACCGTCTCGATGAATTCGCTGACCACTCGAAGCAAGACT 214  
DB 2135206 CTGAATAGCGCATGAGAAAGACTCAAGAAAGTCCATCATGAGCTTCATTAAGCTCT 2135265  
QY 215 T 215  
DB 2135266 T 2135266

## RESULT 7

US-09-974-300-6756/c  
Sequence 6756, Application US/0974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Ib Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085,500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05

;; PRIOR APPLICATION NUMBER: 09/680,598  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/279,526  
;; PRIOR FILING DATE: 2001-03-27  
;; NUMBER OF SEQ ID NOS: 8481  
;; SOFTWARE: FASTSEQ for Windows Version 4.0  
;; SEQ ID NO: 6756  
;; LENGTH: 672  
;; TYPE: DNA  
;; ORGANISM: Bacillus clausii  
US-09-974-300-6756

Query Match 9.9%; Score 59.4; DB 3; Length 672;  
Best Local Similarity 57.1%; Pred. No. 1.8e-07;  
Matches 108; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 33 CAGCACTGCTGCACGATCTGCCCCGGTAGAGGTGTGCTTGAGCGCGTGGCCCGCA 92  
DB 446 CATATCTGCTGCACAAATTTGACCGCTGGAACAAATCGCGGTTTTTAAGTCAACAACT 387  
QY 93 CGGCGAAGCTGCGACCGGAGGCGACCCCGCTGAGACGACGAGCGGTGTAGCC 152  
DB 386 CGCACAAAGCTACGACCAATGTTGACGCCACCTGTGTGTTAAAGCAACCGAGAGCC 327  
QY 153 GTCTGCGACGATCCGACCGCTCTCGATGAAATTCGCGTGAACAGCTCGAA 212  
DB 326 GTACTTATGCGCATAGCACCGTCTCAATTAATGCGCTTGAGAGACGAAAGAAA 267  
QY 213 CTTTGTGCT 221  
DB 266 TTATATCT 258

## RESULT 8

US-10-450-763-18200/C  
;; Sequence 18200, Application US/10450763  
;; Publication No. US20050196754A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
;; FILE REFERENCE: 790CIP3/US  
;; CURRENT APPLICATION NUMBER: US/10/450,763  
;; PRIOR FILING DATE: 2003-06-11  
;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO: 18200  
;; LENGTH: 2493  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SIMILAR  
;; LOCATION: (403)..(1233)  
;; OTHER INFORMATION: 95% homologous to Escherichia coli similar to, accession  
US-10-450-763-18200

Query Match 9.4%; Score 56.4; DB 9; Length 2493;  
Best Local Similarity 59.3%; Pred. No. 1.6e-06;  
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 36 CACCTGTCGACGATCTGCCCCGGTAGAGGTGTGCTTGAGCGCGTGGCCCGACGC 95  
DB 660 CACCTGCGCATATTTCCGAAACCGCAGGTGTGCTTAAAGCCCTGTGGCGGTGA 601  
QY 96 GCAGAACTGCGACCGGAGGCGACCCCGCTGAGACGACGAGCGGTGTAGCCGTC 155  
DB 600 ACAGATTATTAACCTCAGCGCACCCCACTGCGAAGAGACGACGAGCGGTGCA 541

QY 156 GTGGCGACGATCCGACCGCTCTGATGAATTCGCTGAC 197  
DB 540 GTCTCCGGATATACCGCTTTCAGCGCGCTGATGCCAAC 499

## RESULT 9

US-10-450-763-25444/C  
;; Sequence 25444, Application US/10450763  
;; Publication No. US20050196754A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
;; FILE REFERENCE: 790CIP3/US  
;; CURRENT APPLICATION NUMBER: US/10/450,763  
;; PRIOR FILING DATE: 2003-06-11  
;; PRIOR APPLICATION NUMBER: PCT/US01/08631  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 09/540,217  
;; PRIOR FILING DATE: 2000-03-31  
;; PRIOR APPLICATION NUMBER: 09/649,167  
;; PRIOR FILING DATE: 2000-08-23  
;; NUMBER OF SEQ ID NOS: 60736  
;; SOFTWARE: Custom  
;; SEQ ID NO: 25444  
;; LENGTH: 2493  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SIMILAR  
;; LOCATION: (403)..(1233)  
;; OTHER INFORMATION: 95% homologous to Escherichia coli similar to, accession  
US-10-450-763-25444

Query Match 9.4%; Score 56.4; DB 9; Length 2493;  
Best Local Similarity 59.3%; Pred. No. 1.6e-06;  
Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 36 CACCTGTCGACGATCTGCCCCGGTAGAGGTGTGCTTGAGCGCGTGGCCCGACGC 95  
DB 660 CACCTGCGCATATTTCCGAAACCGCAGGTGTGCTTAAAGCCCTGTGGCGGTGA 601  
QY 96 GCAGAACTGCGACCGGAGGCGACCCCGCTGAGACGACGAGCGGTGTAGCCGTC 155  
DB 600 ACAGATTATTAACCTCAGCGCACCCCACTGCGAAGAGACGACGAGCGGTGCA 541  
QY 156 GTGGCGACGATCCGACCGCTCTGATGAATTCGCTGAC 197  
DB 540 GTCTCCGGATATACCGCTTTCAGCGCGCTGATGCCAAC 499

## RESULT 10

US-09-974-300-2461/C  
;; Sequence 2461, Application US/09974300  
;; Patent No. US20020146721A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Berka, Randy M.  
;; TITLE OF INVENTION: Methods for Monitoring Multiple Gene

;; FILE REFERENCE: 10085.500-US  
;; CURRENT APPLICATION NUMBER: US/09/974,300  
;; PRIOR FILING DATE: 2001-10-05  
;; PRIOR APPLICATION NUMBER: 09/680,598  
;; PRIOR FILING DATE: 2000-10-06  
;; PRIOR APPLICATION NUMBER: 60/279,526  
;; PRIOR FILING DATE: 2001-03-27  
;; NUMBER OF SEQ ID NOS: 8481  
;; SOFTWARE: FASTSEQ for Windows Version 4.0  
;; SEQ ID NO: 2461  
;; LENGTH: 594  
;; TYPE: DNA

ORGANISM: *Bacillus licheniformis*  
US-09-974-300-2461

Query Match	9.3%	Score 55.6;	DB 3;	Length 594;	...
Best Local Similarity	58.4%;	Pred. No. 2.4e-06;			
Matches 97; Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0	

QY	56	CCGGTACGACAGTGTCCCTTTGAGCCCGTCGCCGCCGACAGAACTGGACACCGCAAG	115
Db	426	CCGCTTTCATGTTCCGCTTTCAGCCGCCAGGGTCATTAAGTCATTAAGTAAGCCGATCC	367
QY	116	CGACCCCGCTATGACGAGACGACAGCGCGGTATGCGTCGAGCGACGAGATCCGACCG	175
Db	366	GGCTCCGACCTGTGTCAAGACACAAACGAATTGCCATTTCGGCCGACATTATACGG	307
QY	176	TCTCGATGAAATTCGCGCTGCACCACTCGAACAGAACTTTGTGCT	221
Db	306	TCTCGATTCGTACCGCTCATGAGCTCGAATTAATAACTTCAATGT	261

## RESULT 11

```

Sequence 1092, Application US/10194163
Publication No. US20020172976A1
GENERAL INFORMATION:
APPLICANT: Rose, Bruce Carter
TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: PasteSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/194,163
FILING DATE: 04-NOV-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Basu, Shantanu
REGISTRATION NUMBER: 43,318
REFERENCE/DOCKET NUMBER: 529282000101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5995
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1092
SEQUENCE CHARACTERISTICS:
LENGTH: 7685 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..7685
SEQUENCE DESCRIPTION: SEQ ID NO: 1092
US-10-194-163-1092
Query Match 9.0%; Score 53.8; DB 5; Length 7685;
Best Local Similarity 55.7%; Pred. No. 1e-05;
Matches 109; Conservative 0; Mismatches 82; Indels 0; Gaps 0

```

QY	37	ACCTGCTCAGATCTCCCGGTAGCAGATCTCGTTGAGGCGCGTGCGGCCGACGCG	96
Db	743	ATCTGTTCAAGATCTCGGCTGCAGAGATTGCCGTTCCATTCCTGTTTGCCGGTCATG	802
QY	97	CAGAACTGGCAGCGAAGCGCACCCCGCTGTAGACGAAACGACGCGGTGTAGCCGTG	156
Db	803	CAGAAAGGCGAGTCTATTTGGAAACGACCTTAGACGAAATACAGGCGGTAGCCCTGTG	862
QY	157	TGGCGACGATCCGACCGTCTGATGAAATTGGCCGTGACGACGCTCGAACAGAACTTT	216
Db	863	CCCTCAGGGAATAGTACAGATTCCACGAAACGCCCTTCTCTACGGGAAAGATACCTTC	922
QY	217	GTCTGT 221	
Db	923	TTGGT 927	

## RESULT 12

```

1 Sequence 2245, Application US/10156761
2 Publication No. US20030119018A1
3
4 GENERAL INFORMATION:
5
6 APPLICANT: OMURA, SATOSHI
7 APPLICANT: IKEDA, HARO
8 APPLICANT: ISHIKAWA, JUN
9 APPLICANT: HORIKAWA, HIROSHI
10 APPLICANT: SHIBA, TADAYOSHI
11 APPLICANT: SAKAKI, YOSHIYUKI
12 APPLICANT: HATTORI, MASAHIRA
13 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
14 FILE REFERENCE: 249-262
15 CURRENT APPLICATION NUMBER: US/10/156,761
16 CURRENT FILING DATE: 2002-05-29
17 PRIOR APPLICATION NUMBER: JP 2001-204089
18 PRIOR FILING DATE: 2001-05-30
19 PRIOR APPLICATION NUMBER: JP 2001-272697
20 PRIOR FILING DATE: 2001-08-02
21 NUMBER OF SEQ ID NOS: 15109
22
23 SEQ ID NO 2245
24
25 LENGTH: 2715
26
27 TYPE: DNA
28
29 ORGANISM: Streptomyces avermitilis
30
31 FEATURES:
32
33 NAME/KEY: CDS
34 LOCATION: (1)..(2715)
35
36 US-10-156-761-2245

```

Query Match	8.4%	Score 50	DB 6	Length 2715
Best Local Similarity	45.8%	Pred. No. 0.00013		
Matches 173; Conservative	0	Mismatches 205	Indels 0	Gaps 0

Qy	8	GCCGTGTTAAGCGCGGAGCCGAGCGCACACCTGTGTGACGATCTGCGCCGTATGACAGT	67
Db	1012	GCTCCTTGGCGTAACGCTGTGAAGAGCGGACCTGTGTCTGTCAACGCGCGGTCAAGGCGCA	95
Qy	68	GTGCTTTAAGGCGCTGCGCGGCCCGACGCGCGAAGATTGGACACGGAAGCGCACCCGCGCT	127
Db	952	GGTACTTTCAGGGTCTGTGTGTGTGATCGGGAAAGATCGCGCGGTGAGACCGAACTCCGCG	89
Qy	128	GAGACGAGACGAGCGCGGTGTAGCGCGTGTGGACGGAATCCGACCGCTCTCGATGAAAT	187
Db	892	AACATGTTGGCGAATGTGTGGCGCGGTGGAGGAGGCTGTGTGGCGCACACCTTCGCGCTTGA	833
Qy	188	TGCGGTTCGACGACTTCGAACAGGAACCTTTGTGTGTGCTTCCCTGTGTGCACTGCGCT	247
Db	832	ACTTCACGAATCTTGCACGACAGCGCGCTTTGCGCAGCATCTTCGTATGTGTAGACACA	773
Qy	248	CGAGCAGGGATTCACAGAGAGTGTGTCAAGTTTCCCGGTATGTGCTTCAAGGTTGTGGGCGAGT	307
Db	772	GGTCCGTGGCGGTCTGTGCCGCGGCTTGAAGCTCAACCGGTGTAGCTTTAAGCCACAGACGCGG	713
Qy	308	CTTTGGCGCTCGGATTAAGCTTGTTCGAAGAATGTGCGCTGCTTCCGTTTGGCGCGACAGC	367

```
Db 712 GATGACATCGAGACCGGCTGCGGAGCATCGGCGCTCGACATGCAACGAGC 653
QY 368 GCTCCGCGAGCTCGAGA 385
Db 652 CCCAGCCGACGACACCGA 635

RESULT 13
US-10-915-740A-17
; Sequence 17, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Hervé
; APPLICANT: Venter, J. Craig
; APPLICANT: Masiugnani, Vega
; APPLICANT: Galeocci, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarfello, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Piazza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT FILING DATE: 2004-08-11
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US/10/915,740A
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 60872
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-915-740A-17

Query Match 8.4%; Score 50; DB 9; Length 60872;
Best Local Similarity 54.3%; Pred. No. 0.00017;
Matches 101; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 36 CACCTGTCGACGATCTCGCCCGGTAGACAGGTGCTTGAGCGCGTGCAGCGCCGAGCC 95
Db 8452 CAATGCCCCGATGATTTTCGGCGAGCAGTCAATTCGCTTGAAGCCCTCCGCGCGTGA 8511
QY 96 GCAAGACTGCGACCGGAGGCGCACCCCGCTGAGACGAGCGAGGGGTGTAGCCGTC 155
Db 8512 ACAAATGTACTTCCAAAGCGCACCGCACTGTGTAAGAAATGCAAGCGGTGCCGATC 8571
QY 156 GTGGCGACGATCCGCGACCGTCTCGATGAATTTGCCGTGACACCACTGGAACGAACTT 215
Db 8572 CGATTCCGGGATGAAGACGCTTTCAACCGCTTGCCTCGACGACATCCAAAGCCATTT 8631
QY 216 TGTGCT 221
Db 8632 TCGAGT 8637

RESULT 14
US-10-915-740A-1068/c
; Sequence 1068, Application US/10915740A
; Publication No. US20050191316A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Hervé
; APPLICANT: Venter, J. Craig
; APPLICANT: Masiugnani, Vega
; APPLICANT: Galeocci, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarfello, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Piazza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Sequences And Methods Of Their Use
; FILE REFERENCE: 002441.00090
; CURRENT FILING DATE: 2004-08-11
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US/10/915,740A
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: USSN 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: USSN 60/132,068
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1068
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1068
; LENGTH: 2242716
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-915-740A-1068

Query Match 8.4%; Score 50; DB 9; Length 2242716;
Best Local Similarity 54.3%; Pred. No. 0.00024;
Matches 101; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
```

```
QY 36 CACCTGTCGACGATCTCGCCCGGTAGACAGGTGCTTGAGCGCGTGCAGCGCCGAGCC 95
Db 1296347 CAATGCCCCGATGATTTTCGGCGAGCAGTCAATTCGCTTGAAGCCCTCCGCGCGTGA 1296288
QY 96 GCAAGACTGCGACCGGAGGCGCACCCCGCTGAGACGAGCGAGGGGTGTAGCCGTC 155
Db 1296287 ACAAATGTACTTCCAAAGCGCACCGCACTGTGTAAGAAATGCAAGCGGTGCCGATC 1296228
QY 156 GTGGCGACGATCCGCGACCGTCTCGATGAATTTGCCGTGACACCACTGGAACGAACTT 215
Db 1296227 CGATTCCGGGATGAAGACGCTTTCAACCGCTTGCCTCGACGACATCCAAAGCCATTT 1296168
QY 216 TGTGCT 221
Db 1296167 TCGAGT 1296162

RESULT 15
US-10-472-928-1455/c
; Sequence 1455, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926MO
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqMan99, version 1.03
; SEQ ID NO 1455
; LENGTH: 1083
; TYPE: DNA
```



ORGANISM: Streptococcus pneumoniae  
US-10-472-928-1455

Query Match 8.2%; Score 48.8; DB 8; Length 1083;  
Best Local Similarity 53.7%; Pred. No. 0.00027;  
Matches 101; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY	28	AGCGCGACGACCTGCTGCAAGATCTCGCCCGGTAGCAGTGTCGCTTGAGGCGCGTCCGG	87
DB	422	ACGAGCATGATTTGCGCTACGATTTCACTTAATGAGTCACTTCTTAAATCAAA	363
QY	88	CCCGAGCGCGAAGACTGGCAACGAGAGCGCGCACCCCGCTGAGACGAGCGAGCGGTG	147
DB	362	CCAGAGGCAAGAGTACACCGATATTAACGCCGACTGAGTGTCAACAGACAGAT	303
QY	148	TAGCCGTCTGGGAGCGGATCCGCACTGCTTCAATGAAATTGCCGTGACCACTCGAC	207
DB	302	AAACCAATAGTTGACGATGAGTACGCTCAATTAACATACCGTGGGCAATTCAAG	243
QY	208	AGGACTT 215	
DB	242	AGATATT 235	

Search completed: April 7, 2006, 04:35:58  
Job time : 435.403 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 7, 2006, 01:33:46 ; Search time 340.371 Seconds  
(without alignments)  
11689.676 Million cell updates/sec

Title: US-10-611-442-2\_COPY\_17000\_17596  
Perfect score: 597  
Sequence: 1 aacacgagcgtgtgacgccc.....caacgacgagcagcgagcgc 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 333246308 residues  
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N\_Geneseq\_21.\*  
2: geneeqn1980s.\*  
3: geneeqn2000s.\*  
4: geneeqn2001as.\*  
5: geneeqn2001bs.\*  
6: geneeqn2002as.\*  
7: geneeqn2002bs.\*  
8: geneeqn2003as.\*  
9: geneeqn2003bs.\*  
10: geneeqn2003cs.\*  
11: geneeqn2003ds.\*  
12: geneeqn2004as.\*  
13: geneeqn2004bs.\*  
14: geneeqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597	100.0	17596	12 AD114148	Ad114148 M. megaloc
2	77.8	13.0	110000	4 AAI99682_31	Continuation (32 o
3	77.8	13.0	110000	4 AAI99683_31	Continuation (32 o
4	77.6	13.0	1080	14 AC173225	Ac173225 M. xanthu
5	77.6	13.0	13751	14 AC164636	Ac164636 M. xanthu
6	75	12.6	792	14 AC166587	Ac166587 M. xanthu
7	75	12.6	6713	14 AC164330	Ac164330 M. xanthu
8	73.8	12.4	349980	6 AB081849	Ab081849 Bifidobac
9	64.8	10.9	456	11 ABD07607	Abd07607 Pseudomon
10	64.8	10.9	855	11 ABD07733	Abd07733 Pseudomon
11	64.8	10.9	1164	11 ABD07815	Abd07815 Pseudomon
12	63.8	10.7	2000	11 AC135016	Ac135016 Rice stre
13	63.8	10.5	1098	5 AAH67179	Aah67179 C. glutam
14	62.6	10.5	1098	8 ACA00654	Aca00654 C. glutam
15	62.6	10.5	349980	5 AAH88531	Aah88531 C. glutam
16	62.6	10.5	349980	5 AAH88530	Aah88530 C. glutam
17	62.2	10.4	2000	11 AC135412	Ac135412 Rice stre
18	61.4	10.3	15377	14 AC164653	Ac164653 M. xanthu

c	20	60.8	10.2	966	14 AC166085	Ac166085 M. xanthu
c	21	59.4	9.9	672	6 ABK79465	Abk79465 Bacillus
c	22	56.4	9.4	1200	4 AAC82655	Aac82655 E. coli Y
c	23	56.4	9.4	2493	5 AAS89640	Aas89640 DNA encod
c	24	56.4	9.4	2493	5 AAS82396	Aas82396 DNA encod
c	25	55.6	9.3	594	6 AAK75170	Abk75170 Bacillus
c	26	55.4	9.3	1194	11 ACH98601	Ach98601 Klebsiell
c	27	50	8.4	60873	3 AAH1469	Aah1469 N. mening
c	28	50	8.4	110000	3 AAH1490_12	Continuation (13 o
c	29	50	8.4	349980	3 AAF21610	Aaf21610 Neisseria
c	30	48.8	8.2	1083	3 ABX06440	Abx06440 S. pneumo
c	31	48.8	8.2	1086	13 ADK43597	Adk43597 Streptoco
c	32	48.8	8.2	1095	13 ADK93568	Adk93568 Novel S.
c	33	48.8	8.2	1095	14 AEA57438	Aea57438 Streptoco
c	34	48.8	8.2	5558	2 AAH52236	Aav52236 Streptoco
c	35	48.8	8.2	110000	10 ABB56454_06	Continuation (7 of
c	36	48.4	8.1	1339	13 ADK4470	Adk4470 Plant ful
c	37	48.4	8.1	1460	13 ADX10681	Adx10681 Plant ful
c	38	48.4	8.1	1832	13 ADX46207	Adx46207 Plant ful
c	39	48.4	8.1	1832	13 ADX46207	Adx46207 Plant ful
c	40	48.2	8.1	1832	13 ADX46207	Adx46207 Plant ful
c	41	48.2	8.1	573	10 ADK54722	Adk54722 Plant DNA
c	42	48.2	8.1	666	10 ADK75121	Adk75121 Rice phyt
c	43	48.2	8.1	666	10 ADK52933	Adk52933 Rice geno
c	44	48.2	8.1	1348	14 ABB66938	Aeb66938 Rice geno
c	45	47.8	8.0	3647	13 ADR08373	Adr08373 Full leng

## ALIGNMENTS

RESULT 1  
ID: AD114148 standard; DNA; 17596 BP.  
AC: AD114148;  
XX  
XX  
DT: 15-APR-2004 (first entry)  
XX  
XX  
DE: M. megalomicea cosmid KOS205-57-2.3B SEQ ID NO:2.  
XX  
XX  
KM: db; polyketide; enzyme; MegR; MegF; MegK; MegCIV; MegCV; MegBVI; MegBIII;  
MegL; MegM; megosamine; megalomycin.  
OS: Micromonospora megalomicea.  
XX  
XX  
PN: WO2004003169-A2.  
XX  
XX  
PD: 08-JAN-2004.  
XX  
XX  
PF: 30-JUN-2003; 2003WO-US020681.  
XX  
XX  
PR: 28-JUN-2002; 2002US-0393016P.  
XX  
XX  
PA: (KOSA-) KOSAN BIOSCIENCES INC.  
XX  
XX  
PI: Hutchinson RC, Katz L, Reid R, Hu Z, Gramajo H;  
XX  
XX  
DR: WPI; 2004-203379/19.  
XX  
XX  
PT: Novel isolated, purified, or recombinant nucleic acid comprising  
PT: polyketide modifying gene, there gene encodes polyketide modifying enzyme  
PT: e.g., MegR, MegK, or MegM enzymes useful for producing modified  
PT: polyketide.  
XX  
XX  
PS: Example 2; SEQ ID NO 2; 51bp; English.  
XX  
XX  
CC: The invention relates to a novel isolated, purified, or recombinant  
CC: nucleic acid (I) comprising a polyketide modifying gene, where the gene  
CC: encodes a polyketide modifying enzyme chosen from MegR, MegF, MegK,  
CC: MegCIV, MegCV, MegBVI, MegBIII, MegL, and MegM enzymes. A method of the  
CC: invention is useful for producing a modified polyketide, which involves  
CC: culturing a recombinant cell comprising the recombinant nucleic acid

under conditions in which the cell expresses a product of a gene encoded by the nucleic acid under conditions in which the unmodified polyketide is present, and producing the modified polyketide. The cell produces megasamine and can attach megasamine to a polyketide, where the cell, its naturally occurring non-recombinant state cannot produce megasamine. The present sequence contains downstream megalomycin modification enzyme genes.

Sequence 17596 BP; 2325 A; 5966 C; 6604 G; 2701 T; 0 U; 0 Other;

Query Match 100.0%; Score 597; DB 12; Length 17596;

Best Local Similarity 100.0%; Pred. No. 1.7e-132;

Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 AACACGAGCTGTGAGCGCCGCGCCGAGCAGCACTGCTGACAGATTCCCGCGGT 60
DB 17000 AACACGAGCTGTGAGCGCCGCGCCGAGCAGCACTGCTGACAGATTCCCGCGGT 17059
OY 61 AGCAGGTGTGCTTGAAGGCGGTGCGGGCCGAGCGCGAGAACTGAGCAAGCGAGCGCAC 120
DB 17060 AGCAGGTGTGCTTGAAGGCGGTGCGGGCCGAGCGCGAGAACTGAGCAAGCGAGCGCAC 17119
OY 121 CCCGCTGAGAGCAGAGCGCAGCGGTGTAGCCGTGCTGGCGAGCGGATCCGACCGCTCTCG 180
DB 17120 CCCGCTGAGAGCAGAGCGCAGCGGTGTAGCCGTGCTGGCGAGCGGATCCGACCGCTCTCG 17179
OY 181 ATGAATTTGCGGTGACCAAGCTCCGAAAGAACTTTGTCTGTGCTTCCCTTGATGCGA 240
DB 17180 ATGAATTTGCGGTGACCAAGCTCCGAAAGAACTTTGTCTGTGCTTCCCTTGATGCGA 17239
OY 241 CTGGGCTCGGAGAGGTGTGACAGAGAGGTGCTGAGTTGCGCGTGTGCTTCAAGCGTGTG 300
DB 17240 CTGGGCTCGGAGAGGTGTGACAGAGAGGTGCTGAGTTGCGCGTGTGCTTCAAGCGTGTG 17299
OY 301 GCCAGTCTTTCGCGTCCGATTAAGCTTGTGCAAGATGTGCGGTCTTCCGTTCCGCG 360
DB 17300 GCCAGTCTTTCGCGTCCGATTAAGCTTGTGCAAGATGTGCGGTCTTCCGTTCCGCG 17359
OY 361 CCGAGCGCGCTCCGCGAGCTCGAGAGCAAGCTGTGAAGACCGACGCGCGAGCGGTGT 420
DB 17360 CCGAGCGCGCTCCGCGAGCTCGAGAGCAAGCTGTGAAGACCGACGCGCGAGCGGTGT 17419
OY 421 CGTGCGAGATGGGTAGAGCCAGCAGCTGGGGCGAAGCTGACATGTGACCAACCTATCA 480
DB 17420 CGTGCGAGATGGGTAGAGCCAGCAGCTGGGGCGAAGCTGACATGTGACCAACCTATCA 17479
OY 481 CGGTGCAAGAGACGTCAATTGCTCAAGTGAACCAAGAGAGCTTGAAGATGAGAGTCT 540
DB 17480 CGGTGCAAGAGACGTCAATTGCTCAAGTGAACCAAGAGAGCTTGAAGATGAGAGTCT 17539
OY 541 CTCGTGTCTTGGCCATATAGCGGTGAGCTGCGCAATTGACAAACCGCGAGCGGCGC 597
DB 17540 CTCGTGTCTTGGCCATATAGCGGTGAGCTGCGCAATTGACAAACCGCGAGCGGCGC 17596
```

RESULT 2  
AI199682\_31  
Continuation (32 of 45) of AI199682 from base 3100001 (Mycobacterium tuberculosis strain  
WP Sequence Split into 45 fragments LOCUS AI199682 Accession AI199682

WP	Fragment Name	Begin	End
WP	AI199682_00	1	110000
WP	AI199682_01	100001	210000
WP	AI199682_02	200001	310000
WP	AI199682_03	300001	410000
WP	AI199682_04	400001	510000
WP	AI199682_05	500001	610000
WP	AI199682_06	600001	710000
WP	AI199682_07	700001	810000
WP	AI199682_08	800001	910000
WP	AI199682_09	900001	1010000
WP	AI199682_10	1000001	1110000
WP	AI199682_11	1100001	1210000
WP	AI199682_12	1200001	1310000

```
WP AI199682_13 130001 1410000
WP AI199682_14 140001 1510000
WP AI199682_15 150001 1610000
WP AI199682_16 160001 1710000
WP AI199682_17 170001 1810000
WP AI199682_18 180001 1910000
WP AI199682_19 190001 2010000
WP AI199682_20 200001 2110000
WP AI199682_21 210001 2210000
WP AI199682_22 220001 2310000
WP AI199682_23 230001 2410000
WP AI199682_24 240001 2510000
WP AI199682_25 250001 2610000
WP AI199682_26 260001 2710000
WP AI199682_27 270001 2810000
WP AI199682_28 280001 2910000
WP AI199682_29 290001 3010000
WP AI199682_30 300001 3110000
WP AI199682_31 310001 3210000
WP AI199682_32 320001 3310000
WP AI199682_33 330001 3410000
WP AI199682_34 340001 3510000
WP AI199682_35 350001 3610000
WP AI199682_36 360001 3710000
WP AI199682_37 370001 3810000
WP AI199682_38 380001 3910000
WP AI199682_39 390001 4010000
WP AI199682_40 400001 4110000
WP AI199682_41 410001 4210000
WP AI199682_42 420001 4310000
WP AI199682_43 430001 4410000
WP AI199682_44 440001 441529
```

Query Match 13.0%; Score 77.8; DB 4; Length 110000;  
Best Local Similarity 61.7%; Pred. No. 1.4e-08;  
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
OY 26 CGAGCGCGAGCACTGTCGACGATCTGCGCCGTTAGCAGGTGTGCTTGAAGCGGTGCG 85
DB 90207 CGCGCGAGCGACCTGCTCGAGAGATCTCCGCGTGAAGTGTGCGATCAATCGGCTT 90266
OY 86 GCGCCGAGCGCGAGCACTGTCGACGCGAAGCGCACCCCGCTTGAAGCAGAGCGCGCG 145
DB 90267 GCGCCGAGCGCGAGCACTGTCGACGCGAAGCGCACCCCGCTTGAAGCAGAGCGCG 90326
OY 146 TGTAGCGCGTGTGCGACGATTCGCAACCGCTGATGAATTTGCGCGTGAACGAGCTGCA 205
DB 90327 TGTAGCGCGTGTGCGACGATTCGCAACCGCTGATGAATTTGCGCGTGAACGAGCTGCA 90386
OY 206 ACAGAACTTTGTGCTGCGC 226
DB 90387 ACAGAACTTTGTGCTGCGC 90407
```

RESULT 3  
AI199683\_31

Continuation (32 of 44) of AI199683 from base 3100001 (Mycobacterium tuberculosis strain  
WP Sequence Split into 44 fragments LOCUS AI199683 Accession AI199683

WP	Fragment Name	Begin	End
WP	AI199683_00	1	110000
WP	AI199683_01	100001	210000
WP	AI199683_02	200001	310000
WP	AI199683_03	300001	410000
WP	AI199683_04	400001	510000
WP	AI199683_05	500001	610000
WP	AI199683_06	600001	710000
WP	AI199683_07	700001	810000
WP	AI199683_08	800001	910000
WP	AI199683_09	900001	1010000
WP	AI199683_10	1000001	1110000
WP	AI199683_11	1100001	1210000
WP	AI199683_12	1200001	1310000
WP	AI199683_13	1300001	1410000

WP AA19683\_14 1400001 1510000  
WP AA19683\_15 1500001 1610000  
WP AA19683\_16 1600001 1710000  
WP AA19683\_17 1700001 1810000  
WP AA19683\_18 1800001 1910000  
WP AA19683\_19 1900001 2010000  
WP AA19683\_20 2000001 2110000  
WP AA19683\_21 2100001 2210000  
WP AA19683\_22 2200001 2310000  
WP AA19683\_23 2300001 2410000  
WP AA19683\_24 2400001 2510000  
WP AA19683\_25 2500001 2610000  
WP AA19683\_26 2600001 2710000  
WP AA19683\_27 2700001 2810000  
WP AA19683\_28 2800001 2910000  
WP AA19683\_29 2900001 3010000  
WP AA19683\_30 3000001 3110000  
WP AA19683\_31 3100001 3210000  
WP AA19683\_32 3200001 3310000  
WP AA19683\_33 3300001 3410000  
WP AA19683\_34 3400001 3510000  
WP AA19683\_35 3500001 3610000  
WP AA19683\_36 3600001 3710000  
WP AA19683\_37 3700001 3810000  
WP AA19683\_38 3800001 3910000  
WP AA19683\_39 3900001 4010000  
WP AA19683\_40 4000001 4110000  
WP AA19683\_41 4100001 4210000  
WP AA19683\_42 4200001 4310000  
WP AA19683\_43 4300001 4403765

Query Match 13.0%; Score 77.8; DB 4; Length 110000;  
Best Local Similarity 61.7%; Pred. No. 1.4e-08;  
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 26 CGAGCGCCGACGCTCTGACGATCTCCCGGTAAGAGTGTCTTGAAGCCCTCCG 85  
DB 84316 CCGCGGACGACGCTCTGAGAGATCTCCGCGGTGATGTTGCGAGTCAATCCGCCCT 84375  
QY 86 GCGCCGACGCGGACGACGCTGCGAGCGGACCGCCGCTTGAAGACGACGCGCG 145  
DB 84376 GCGCAGTGTGACAGAGAGCGGACGATCCCGGCGGCTTGCAGGAATGCAAGCCG 84435  
QY 146 TGTAGCGCTGTGTGGCGAGATCCGACCGCTCTGATGAATTTGCGGTGACCGCTCGA 205  
DB 84436 TTTTGGCGCGCGGATAGCGATTCAGACCGATTGCAATGTTACCGTGAACGCGCCGCC 84495  
QY 206 ACAGAACTTTGTCTGTGCG 226  
DB 84496 ACAACGTCTTTCGAGTCTGCG 84516

RESULT 4  
ACL73225/c  
ID ACL73225 standard; DNA; 1080 BP.

XX AC ACL73225;  
XX DT 02-JUN-2005 (first entry)  
XX DE M. xanthus gene sequence, seq id 9688.  
XX KM Transgenic plant; DNA replication; gene regulation; gene expression;  
XX KM gene; db.  
XX OS Myxococcus xanthus.  
XX PN US6833447-B1.  
XX PD 21-DEC-2004.  
XX PF 10-JUL-2001; 2001US-00902540.  
XX

PR 10-JUL-2000; 2000US-0217883P.  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
PA Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
XX WPI; 2005-028716/03.  
DR

PT New substantially purified Myxococcus xanthus nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.

Example 2; SEQ ID NO 9688; 25pp; English.

CC The invention relates to a substantially purified nucleic acid molecule  
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC set of interest. Sequences given in records for SEQ IDs 1850-9691 represent a  
CC Myxococcus xanthus. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from USPRO

Sequence 1080 BP; 173 A; 403 C; 343 G; 161 T; 0 U; 0 Other;

Query Match 13.0%; Score 77.6; DB 14; Length 1080;  
Best Local Similarity 61.3%; Pred. No. 8.3e-09;  
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 34 AGCACTGTGACGATCTGCGCCCGTACAGGTGTGCTTGAAGCCGTGCGCCCGAC 93  
DB 431 AGTATGTGCCCGACGATTTGCGCGGCGTCAAGTGTGCGAGAGCCCATCTGCCGGTG 372  
QY 94 GCGCAGAACTGCGACGCGGACCGACCCCGCTTGAAGACGACGAGCGGTGTGCGG 153  
DB 371 GCGCAGAAAGCGACCGCCATGCGCGGCGCTGTGTGCTATGCAACCGTGGCGCG 312  
QY 154 TCGTGGCGAGGATCGCACCGCTCTGATGAATTTGCGTGAACCGACTCGAAGAAC 213  
DB 311 CCTTGAACCGATGACGAGTTTCAATGATGTGCTCCGCTGATGCGGACGACGACG 252  
QY 214 TTTGTGCTGTGCTTCCCTGTG 237  
DB 251 TTGTGCGTGAAGCCGTGCTGCTG 228

RESULT 5  
ACL64636/c  
ID ACL64636 standard; DNA; 13751 BP.

XX AC ACL64636;  
XX DT 02-JUN-2005 (first entry)  
XX DE M. xanthus DNA fragment, seq id 1099.  
XX KM Transgenic plant; DNA replication; gene regulation; gene expression; db.  
XX KM Myxococcus xanthus.  
XX PN US6833447-B1.  
XX PD 21-DEC-2004.  
XX PF 10-JUL-2001; 2001US-00902540.  
XX PR 10-JUL-2000; 2000US-0217883P.  
XX (MONS ) MONSANTO TECHNOLOGY LLC.

XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
PI WPI; 2005-028716/03.  
XX  
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.  
XX  
XX Example 1; SEQ ID NO 1099; 25bp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a  
CC set of 1849 contig and singleton sequences comprising coding sequences,  
CC DNA replication elements, promoters and other regulatory elements from  
CC the genome of the bacterium *Myxococcus xanthus*. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO  
XX  
SQ Sequence 13751 BP; 2088 A; 5011 C; 4651 G; 2001 T; 0 U; 0 Other;  
Query Match 13.0%; Score 77.6; DB 14; Length 13751;  
Best Local Similarity 61.3%; Pred. No. 1.2e-08;  
Matches 125; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
XX  
QY 34 AGCAGCTGCTGACAGATCTGCGCGGTGAGAGTGTGCTTGAAGACCGTCGCGACCGAC 93  
DB 5590 AGTATCTGCCCCAGATTTCGCCGCGCTCAGGTGCGTGAAGACCCCATCTGCCCGGTG 5511  
QY 94 GCGCAGAACTGCGACGCGAAGCGCGACCCCGCTGAGAGAGACGCGAGCGGTGTAGCCG 153  
DB 5530 GCGCAGAAAGACGAGCCCATGCGCGACCGCGCTGCGTGAATGACACCGTGGCGCG 5471  
QY 154 TCGTGGCGACGAGATCCGACCGCTCTCATGAAATTTCCGTGACCACTCGAAGACGAC 213  
DB 5470 CCGTTGAACCGCATTCAGACCGTTCATGATGTGCGCGTGTCCAGTCCAGACGACG 5411  
QY 214 TTTGTGCTGTGCGCTTCCCTGTGATG 237  
DB 5410 TTGTGCGTGAAGCGGTGCTGTG 5387

RESULT 6  
ACL6587/C  
ID ACL6587 standard; DNA; 792 BP.  
XX  
XX ACL6587;  
AC  
XX  
XX 02-JUN-2005 (first entry)  
DT  
XX  
XX M. xanthus gene sequence, seq id 3050.  
DE  
XX  
XX Transgenic plant; DNA replication; gene regulation; gene expression;  
KW gene; ds.  
XX  
XX *Myxococcus xanthus*.  
OS  
XX  
XX US6833447-B1.  
PN  
XX  
XX 21-DEC-2004.  
PD  
XX  
XX 10-JUL-2001; 2001US-00902540.  
PF  
XX  
XX 10-JUL-2000; 2000US-0217883P.  
PR  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
PA  
XX  
XX

PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
XX WPI; 2005-028716/03.  
XX  
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.  
XX  
XX Example 2; SEQ ID NO 3050; 25bp; English.  
XX  
XX The invention relates to a substantially purified nucleic acid molecule  
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC of interest. Sequences given in records for SEQ IDs 1850-9691 represent a  
CC set of about 7842 genes or partial genes from the genome of the bacterium  
CC *Myxococcus xanthus*. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from USPTO  
XX  
XX  
SQ Sequence 792 BP; 148 A; 267 C; 257 G; 120 T; 0 U; 0 Other;  
Query Match 12.6%; Score 75; DB 14; Length 792;  
Best Local Similarity 58.9%; Pred. No. 3.3e-08;  
Matches 129; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
XX  
QY 3 CACGAGCCTTTTGAACGCGCGGCCCGGACGCGCAGACCTGCTGACCATCTGCCCGGTAG 62  
DB 459 CTCCTTTTTCGAAACCTCGCGGTTCACCGGTGACCTTGAGCCACATCTCGTGGAGGT 400  
QY 63 CAGGTGTGCTTGAAGCGCGTTCGCGCGGACGCGCAGAACTGCGACCGGAGCGCACCC 122  
DB 399 CAGGTGTGCTTGAAGCGCGTTCGCGCGGACGCGCAGAACTGCGACCGGAGCGCACCC 340  
QY 123 CGCCTGAGACGAGACGCGAGCGGTGTAGCCGTGCTGCGAAGATCCGACCGTCTGAT 182  
DB 339 CACCTGTGTGAGACGCGAACAAGTCCTGCGGTCTCGGTGGGATGTAGACGATTCGAT 280  
QY 183 GAAATTCGCTGACGACGCTGGAACGAGAACTTTGTGCT 221  
DB 279 GTACGCGCGGTCCGCGCTTCCACCGGTACTTGTATGCT 241

RESULT 7  
ACL64330  
ID ACL64330 standard; DNA; 6713 BP.  
XX  
XX ACL64330;  
AC  
XX  
XX 02-JUN-2005 (first entry)  
DT  
XX  
XX M. xanthus DNA fragment, seq id 793.  
DE  
XX  
XX Transgenic plant; DNA replication; gene regulation; gene expression; ds.  
KW  
XX  
XX *Myxococcus xanthus*.  
OS  
XX  
XX US6833447-B1.  
PN  
XX  
XX 21-DEC-2004.  
PD  
XX  
XX 10-JUL-2001; 2001US-00902540.  
PF  
XX  
XX 10-JUL-2000; 2000US-0217883P.  
PR  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
PA  
XX  
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
PI WPI; 2005-028716/03.  
XX  
XX

XX New substantially purified *Myxococcus xanthus* nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing  
PT mutations in a gene of interest.  
XX  
PS Example 1; SEQ ID NO 793; 25pp; English.  
XX  
CC The invention relates to a substantially purified nucleic acid molecule  
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.  
CC The nucleic acid is useful for determining gene expression, identifying  
CC mutations in a gene of interest, and for constructing mutations in a gene  
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a  
CC set of 1849 contigs and singleton sequences comprising coding sequences,  
CC DNA replication elements, promoters and other regulatory elements from  
CC the genome of the bacterium *Myxococcus xanthus*. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO  
XX  
SQ Sequence 6713 BP; 984 A; 2232 C; 2234 G; 1263 T; 0 U; 0 Other;  
Query Match 12.6%; Score 75; DB 14; Length 6713;  
Best Local Similarity 58.9%; Pred. No. 4,4e-08;  
Matches 129; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
QY 3 CAGAGCGCTTGAACCGCGGCGCCGAGCGCCAGACCTGCTGCACGATTCGCGCGGTAG 62  
Db 336 CTGTGTTCTTGAACCTGCGGCTTTCACCGCGGTGACCTGCGCCCACTCTCGTGGGGGT 395  
QY 63 CAGGTGCGCTTGAAGCGCGCTGCGCGCCGAGCGCGACAGTGGCAGCGGAGGCGCACCC 122  
Db 396 CAGGTGCGCTTGAAGCGCGCTGCGCGCCGAGCGCGACAGTGGCAGCGGAGGCGCACCC 455  
QY 123 CGCTGAGACGAGACGCGAGCGGTGTAGCCGCTGTGCGACGATCCGACCGCTTCGAT 182  
Db 456 CACCTGCGTGGACACGACGACAACTGCTGCGGTCTGCGTGGGATGTAGACGATTCGAT 515  
QY 183 GAATTGCGCTGACACGCTCGAAGCAAGAACTTGTCT 221  
Db 516 GTAGCGCGCTGCGCTGCTTTCACCGGTACTGATGT 554  
RESULT 8  
ID ABQ81849 standard; DNA; 349980 BP.  
XX  
AC ABQ81849;  
XX  
DT 19-NOV-2002 (first entry)  
XX  
DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1105.  
XX  
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KM antidiarrheic; antibiotic; inhibitor of *Salmonella*; detection;  
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
KM rotavirus; food composition; pharmaceutical composition; gene; ds.  
XX  
OS Bifidobacterium longum.  
OS Synthetic.  
XX  
PN EP1227152-A1.  
XX  
PD 31-UTL-2002.  
XX  
PF 30-JAN-2001; 2001EP-00102050.  
XX  
PR 30-JAN-2001; 2001EP-00102050.  
XX  
PA (NEST) SOC PROD NESTLE SA.  
XX  
DR WPI; 2002-668397/72.

XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
PT in a biological sample.  
XX  
PS Disclosure; SEQ ID NO 1105; 80pp; English.  
XX  
CC The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
CC least 90% identity or which hybridises with the sequences given in  
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
CC fusion protein, comprising a sequence selected from 1097 sequences given  
CC in ABQ65258 to ABQ65354 ligated in frame to a polynucleotide encoding a  
CC heterologous polypeptide. (I) has antidiarrheic and antibacterial  
CC activities, and can be used as an inhibitor of *Salmonella*. (I) (which is  
CC a probe) is useful for the detection and/or identification of  
CC Bifidobacterium longum in a biological sample. A carrier containing the  
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be  
CC used for preventing and/or treating diarrhoea brought about by pathogenic  
CC bacteria and/or rotavirus. The carrier is a food composition selected  
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
CC products, ice-creams, fermented cereal based products, milk based  
CC powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent  
CC Bifidobacterium related nucleotide sequences given in the sequence  
CC listing from the present invention but not mentioned further within the  
CC specification. N.B. The sequence data for this patent is not represented  
CC in the printed specification but is based on sequence information  
XX  
SQ Sequence 349980 BP; 68046 A; 106490 C; 106389 G; 69055 T; 0 U; 0 Other;  
Query Match 12.4%; Score 73.8; DB 6; Length 349980;  
Best Local Similarity 60.0%; Pred. No. 1.5e-07;  
Matches 123; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 26 CGAGCGCCAGACCTGCTGCACGATCTCGCCCGGTGACGAGTGTGCTTGAAGCGCGC 85  
Db 85216 CGGCCACAGCGACCTTTCGATTTATTTACCGGTGACATGTTACGGGTGACCAAGCT 85157  
QY 86 CGCCGACGCGGCAAACTGACGCGAGCGGACCGCCGCTTGAAGACGACGAGCGCG 145  
Db 85156 TGCCGCGGTGCGGACGAGAGCGGAGTCCATGCGCGACGCGACCTGCGAGGATGACAGCG 85097  
QY 146 TGTAGCGCTGCTGCGGACGAGTCCGACCGCTCTGATGAAATTCGCTGACGACGCTCGA 205  
Db 85096 TGTGCGGCTGCGGATTCGATTAAGCAAGATTCGATGAGGAGCGCGTGAACAGCTTCC 85037  
QY 206 ACAAGAACTTGTGCTGCTGCTTCC 230  
Db 85036 ACAAGCTTGTGATGTGTGCTTCC 85012  
RESULT 9  
ID ABD07607 standard; DNA; 456 BP.  
XX  
AC ABD07607;  
XX  
DT 29-UTL-2004 (first entry)  
XX  
DE *Pseudomonas aeruginosa* polynucleotide #6211.  
XX  
KM Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
KM antibacterial.  
XX  
OS *Pseudomonas aeruginosa*.  
OS  
XX  
PN US6551795-B1.



XX PD 22-APR-2003.  
XX PF 18-FEB-1999; 99US-00252991.  
XX PR 18-FEB-1998; 98US-0074786P.  
XX PR 27-JUL-1998; 98US-0094190P.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Rubinfeld MJ, Nolling J, Deloughery C, Bush D;  
XX DR WPI; 2003-615309/58.  
XX DR P-PSDB; ABO74036.  
XX PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX PT pathological conditions resulting from bacterial infection.  
XX PS Disclosure; SEQ ID NO 6211; 455bp; English.  
XX CC The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
XX CC polynucleotides encoding them. The sequences are useful in diagnosis and  
XX CC therapy of pathological conditions, as molecular targets for diagnostics,  
XX CC prophylaxis and treatment of pathological conditions resulting from a  
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,  
XX CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
XX CC effective antibacterial targets, as targets for antibacterial drugs,  
XX CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
XX CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
XX CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
XX CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
XX CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
XX CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
XX CC The sequence data for this patent did not form part of the printed  
XX CC specification but was obtained in electronic format from USPTO at  
XX CC seqdata.uspto.gov/sequence.html  
XX SQ Sequence 456 BP; 82 A; 142 C; 147 G; 85 T; 0 U; 0 Other;  
Query Match 10.9%; Score 64.8; DB 11; Length 456;  
Best Local Similarity 60.0%; Pred. No. 8.3e-06;  
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 36 CACCTGTGACGATCTGCGCCGGTAGAGGTGTGCTTGAAGCCGTCGCGCCGACGC 95  
DB 144 CACCTGGCCGATCACTCGGGGGGGGTAGGTGCTGTGAAACCTCTTCCCGGTGGA 203  
QY 96 GCAGAACTGGACGCGGAAAGGCGCACCCCGCTGAGACGAGCGCGGTGTAGCCGTC 155  
DB 204 GCAGAACTGGCAATCCAGGGGCGAGCCCGCTGGGAGGACACGACAGGGTGGCCGTC 263  
QY 156 GTGGCGACGATCCGCAACCGTCTGATGAATTTGCCCTGACACGACTGGAACAGAACTT 215  
DB 264 GCCCTGGGGAGTGTACACGCTTCGACGCAACTGCCGAGCGCACCCCGGACCACTT 323

RESULT 10  
ABD07733/C  
ID ABD07733 standard; DNA; 855 BP.  
XX AC ABD07733;  
XX DT 29-JUL-2004 (first entry)  
XX DE *Pseudomonas aeruginosa* polynucleotide #6337.  
XX KW Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
XX KW antibacterial.  
XX OS *Pseudomonas aeruginosa*.  
XX PN US651795-B1.

XX PD 22-APR-2003.  
XX PF 18-FEB-1999; 99US-00252991.  
XX PR 18-FEB-1998; 98US-0074786P.  
XX PR 27-JUL-1998; 98US-0094190P.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Rubinfeld MJ, Nolling J, Deloughery C, Bush D;  
XX DR WPI; 2003-615309/58.  
XX DR P-PSDB; ABO74162.  
XX PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX PT pathological conditions resulting from bacterial infection.  
XX PS Disclosure; SEQ ID NO 6337; 455bp; English.  
XX CC The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
XX CC polynucleotides encoding them. The sequences are useful in diagnosis and  
XX CC therapy of pathological conditions, as molecular targets for diagnostics,  
XX CC prophylaxis and treatment of pathological conditions resulting from a  
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,  
XX CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
XX CC effective antibacterial targets, as targets for antibacterial drugs,  
XX CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
XX CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
XX CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
XX CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
XX CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
XX CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
XX CC The sequence data for this patent did not form part of the printed  
XX CC specification but was obtained in electronic format from USPTO at  
XX CC seqdata.uspto.gov/sequence.html  
XX SQ Sequence 855 BP; 168 A; 270 C; 263 G; 154 T; 0 U; 0 Other;  
Query Match 10.9%; Score 64.8; DB 11; Length 855;  
Best Local Similarity 60.0%; Pred. No. 9.1e-06;  
Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 36 CACCTGTGACGATCTGCGCCGGTAGAGGTGTGCTTGAAGCCGTCGCGCCGACGC 95  
DB 436 CACCTGGCCGATCACTCGGGGGGGGTAGGTGCTGTGAAACCTCTTCCCGGTGGA 377  
QY 96 GCAGAACTGGACGCGGAAAGGCGCACCCCGCTGAGACGAGCGCGGTGTAGCCGTC 155  
DB 376 GCAGAACTGGCAATCCAGGGGCGAGCCCGCTGGGAGGACACGACAGGGTGGCCGTC 317  
QY 156 GTGGCGACGATCCGCAACCGTCTGATGAATTTGCCCTGACACGACTGGAACAGAACTT 215  
DB 316 GCCCTGGGGAGTGTACACGCTTCGACGCAACTGCCGAGCGCACCCCGGACCACTT 257

RESULT 11  
ABD07815/C  
ID ABD07815 standard; DNA; 1164 BP.  
XX AC ABD07815;  
XX DT 29-JUL-2004 (first entry)  
XX DE *Pseudomonas aeruginosa* polynucleotide #6419.  
XX KW Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
XX KW antibacterial.  
XX OS *Pseudomonas aeruginosa*.  
XX PN US651795-B1.

PD	22-APR-2003.	
XD		
PD	18-FEB-1999;	99US-00252991.
XX		
PR	18-FEB-1998;	98US-0074788P.
PR	27-JUL-1996;	98US-0094190P.
XX		
PA	(GENO-) GENOME THERAPEUTICS CORP.	
XX		
PI	Rubenfield MJ, Nolling J, DeLoughery C, Bush D;	
DR	WPI: 2003-615309/58.	
DR	P-FSDB; ABO74244.	
XX		
PT	Novel isolated nucleic acid encoding <i>Pseudomonas aeruginosa</i> polypeptide, useful as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from bacterial infection.	
PT		
XX		
PS	Disclosure; SEQ ID NO 6419; 455bp; English.	
XX		
CC	The invention relates to <i>Pseudomonas aeruginosa</i> polypeptides and the polynucleotides encoding them. The sequences are useful in diagnosis and therapy of pathological conditions, as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from a bacterial infection, for evaluating a compound, such as a polypeptide, for the ability to bind a <i>P. aeruginosa</i> nucleic acid, as components of effective antibacterial targets, as targets for antibacterial drugs, including anti- <i>P. aeruginosa</i> drugs, as templates for recombinant production of <i>P. aeruginosa</i> -derived peptides or polypeptides, as target components for diagnosis and/or treatment of <i>P. aeruginosa</i> -caused infection, and in detection of <i>P. aeruginosa</i> sequences or other sequences of <i>Pseudomonas</i> species using biochip technology. Sequences ABD01397-ABD17667 represent <i>P. aeruginosa</i> polynucleotides of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at <a href="http://seqdata.uspto.gov/sequence.html">seqdata.uspto.gov/sequence.html</a>	
CC		
XX		
SQ	Sequence 1164 BP; 234 A; 361 C; 356 G; 213 T; 0 U; 0 Other;	
	Query Match 10.9%; Score 64.8; DB 11; Length 1164;	
	Best Local Similarity 60.0%; Pred. No. 9,5e-06;	
	Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;	
QY	36 CACCTGCTGCAGATCTCGCCCGGTAGCAGGTGTGCTTGAGGCCGTGCGGCGCCGACGC 95	
DB	456 CACCTGGCCGATCACTCTGGGCGCGGTAGAGTCTGTGTGAACCTCTTGCCGGTGA 397	
QY	96 GCAGAACTGCGACCGGAGAGCGCACCCCGCTTGACGAGACGACGAGCGGTGTAGCCGTC 155	
DB	396 GCAGAACTGCAATCAAGGCGGACCGCGCTGGGAGGACACGACAGGGTGGCGCGTCC 337	
QY	156 GTGGCGACGATCCGACACCGTCTCGATTGAATTGCCGTTCGACACCACTGGAACGAACTT 215	
DB	336 GCCCTGGGGATGTACACGCTTCGACCAACTGCCGAGCGCAACCGACCAACCCACTT 277	
RESULT 12		
ABD07529		
ID	ABD07529	standard; DNA; 1764 BP.
XX		
AC	ABD07529;	
XX		
DT	29-JUL-2004	(first entry)
XX		
DE	<i>Pseudomonas aeruginosa</i> polynucleotide #6133.	
XX		
KW	Bacterial infection; gene; ds; <i>Pseudomonas aeruginosa</i> infection;	
KW	antibacterial.	
XX		
OS	<i>Pseudomonas aeruginosa</i> .	
XX		
PN	US6551795-B1	

XX	FD	22-APR-2003.
XX	PF	18-FEB-1999; 99US-00252991.
XX	PR	18-FEB-1998; 98US-0074788P.
XX	PR	27-JUL-1998; 98US-0094190P.
XX	PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	PI	Rubenfield M <sup>c</sup> , Nolling J, Deloughery C, Bush D;
XX	DR	WPI, 2003-615309/58.
XX	P	P-PsDB; AEO73958.
PT	PT	Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT	PT	useful as molecular targets for diagnostics, prophylaxis and treatment of
PT	PT	pathological conditions resulting from bacterial infection.
XX	PS	Dicloesure; SEQ ID NO 6133; 455pp; English.
CC	CC	The invention relates to Pseudomonas aeruginosa polypeptides and the
CC	CC	polynucleotides encoding them. The sequences are useful in diagnosis and
CC	CC	therapy of pathological conditions, as molecular targets for diagnostics,
CC	CC	prophylaxis and treatment of pathological conditions resulting from a
CC	CC	bacterial infection, for evaluating a compound, such as a polypeptide,
CC	CC	for the ability to bind a P. aeruginosa nucleic acid, as components of
CC	CC	effective antibacterial targets, as targets for antibacterial drugs,
CC	CC	including anti-P. aeruginosa drugs, as templates for recombinant
CC	CC	production of P. aeruginosa-derived peptides or polypeptides, as target
CC	CC	components for diagnosis and/or treatment of P. aeruginosa-caused
CC	CC	infection, and in detection of P. aeruginosa sequences or other sequences
CC	CC	of Pseudomonas species using bioclip technology. Sequences ABD01397-
CC	CC	ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
CC	CC	The sequence data for this patent did not form part of the printed
CC	CC	specification but was obtained in electronic format from USPTO at
CC	CC	seqdata.uspto.gov/sequence.html
SQ	SQ	Sequence 1764 BP; 308 A; 544 C; 557 G; 355 T; 0 U; 0 Other;
		Query Match 10.9%; Score 64.8; DB 11; Length 1764;
		Best Local Similarity 60.0%; Pred. No. 1e-05;
		Matches 108; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
OY		36 CACCTGTGCAGCATCTCGCCGGGTACGAGGTTCGCTTGAGGCGCTCGCGGCCGAAGC 95
Db		1525 CACCTGGCCGATCACTCTCGCGGGGTGAGGTCCCTGTGAACCCTTGTCCGGTSGA 1584
OY		96 GCAGAAGTGGCACGCGAAGGCGCACCCCGCTGAGACGACGCGAGGCGGTGTAGCCGTC 155
Db		1585 GCAGAAATGTGAATTCAGGGGGCGAGCCCGCTGGAGAGACACGACAGGGGTGCCGCTCC 1644
OY		156 GTGCGCAGCGATCCGCAACCGTTCTGATGAATAATTGCCGTCGACACAGCTCGAACAAGAACTT 215
Db		1645 GCCCTGGGAGTAGTACACGCTCTGACGCAATGCCCCGAGACGCAACCGGACCACTT 1704
RESULT 13		
ACL35016		
ID	ACL35016	standard; cDNA; 2000 BP.
XX	AC	
XX	AC	ACL35016;
XX	DT	
XX	DT	02-JUN-2005 (first entry)
DE	XX	Rice stress-regulated promoter SEQ ID NO:13579.
XX	XX	
KW	ss;	abiotic stress tolerance; transgenic plant; plant; cereal;
XX	agiculture.	
OS	Oryza sativa.	
PN	WO2003008540-A2.	

XX 30-JAN-2003.  
PD  
XX  
XX 21-JUN-2002; 2002MO-US019668.  
PF  
XX 22-JUN-2001; 2001US-0300112P.  
PR 24-AUG-2001; 2001US-0314662P.  
PR 26-SEP-2001; 2001US-0325277P.  
PR 21-NOV-2001; 2001US-0332132P.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Kieps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F,  
PI Moughamer T, Provart N, Ricke D, Zhu T;  
XX  
DR WPI; 2003-248011/24.  
XX  
PT New stress-responsive nucleic acid, useful for altering the  
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold  
PT stress, salt stress or osmotic stress.  
XX  
PS Claim 48; SEQ ID NO 13579; 89pp; English.  
XX  
CC The invention relates to novel abiotic stress responsive polynucleotides  
CC and polypeptides. Also disclosed are vectors, expression cassettes, host  
CC cells, and plants containing such polynucleotides. Also disclosed are  
CC methods for using the polynucleotides and polypeptides to alter the  
CC responsiveness of a plant to abiotic stress. The invention is useful in  
CC agriculture. The nucleic acid is useful for determining whether a test  
CC plant has been exposed to an abiotic stress condition. It is also useful  
CC for selecting an agent that alters abiotic stress regulated  
CC polynucleotide expression in a plant cell, and to identify a homolog or  
CC molecule to an abiotic stress responsive polynucleotide. The nucleic acid  
CC and the polypeptide encoded by it are useful in altering the  
CC stress responsiveness of a plant to an abiotic stress, such as cold stress, salt  
CC stress, osmotic stress or any of their combinations. The present sequence  
CC is used in the exemplification of the invention  
XX  
SQ Sequence 2000 BP; 342 A; 599 C; 666 G; 389 T; 0 U; 4 Other;  
Query Match 10.7%; Score 63.8; DB 11; Length 2000;  
Best Local Similarity 55.1%; Pred. No. 1.8e-05;  
Matches 125; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 11 TGTGAGCGCGGCGCCGAGACCTGCTGACATCTCGCCGCTGAGCTGTC 70  
DB 711 TGCCAGGTGGCGTGGCGCAACCACTGCGCATATCTCGCGGTGTCAAGTTGC 770  
QY 71 GCTTGAAGCCGTGCGGCGCCGACGCGAGAACTGGACGCGAGCGCACCCCGCTGAG 130  
DB 771 GATTGAAGCCCTTGAGTGGCGGTGAGAGCAAGAGTGCAGTTCAAGGCCGACCGACTGCG 830  
QY 131 ACGAGACGCAAGCGGTGTAGCCGCTGTGCGACGAGATCCGACCGCTCTCGATGAATTGCG 190  
DB 831 ACGACACGCGACAGCGGTGCGCGGCTTGTCTCGGATGTACACGCTCTCGATGAGGCTTCT 890  
QY 191 CGTGACACGAGCTCGAAGAGAACTTGTGCTGTGCTTCCCGTGTG 237  
DB 891 TGCATTCACGCGCATTCGCGACAGACCACTTGTGTGCTCGCTCGCG 937

RESULT 14  
AAH67179/c  
ID AAH67179 standard; DNA; 1098 BP.

XX AAH67179;  
AC  
XX 26-SEP-2001 (first entry)  
DT  
XX C glutamicum coding sequence fragment SEQ ID NO: 2214.  
DE  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.

XX  
OS Corynebacterium glutamicum.  
XX  
XX EPI108790-A2.  
FN  
XX 20-JUN-2001.  
PD  
XX 18-DEC-2000; 2000EP-00127688.  
PF  
XX 16-DEC-1999; 99JP-00377484.  
PR 07-APR-2000; 2000JP-00159162.  
PR 03-AUG-2000; 2000JP-00280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
PI  
DR WPI; 2001-376931/40.  
DR P-PSDB; AAG91960.  
XX  
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analyzing  
PT expression profile or pattern of a gene and identifying homologous gene.  
XX  
PS Claim 8; SEQ ID NO 2214; 246pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and analyzing  
CC the expression profile or expression pattern of a gene derived from  
CC Corynebacterium bacterium, and identifying a homolog of a gene derived from  
CC Corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino  
CC acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the European Patent Office  
XX  
SQ Sequence 1098 BP; 221 A; 295 C; 304 G; 278 T; 0 U; 0 Other;  
Query Match 10.5%; Score 62.6; DB 5; Length 1098;  
Best Local Similarity 59.1%; Pred. No. 3.1e-05;  
Matches 107; Conservative 0; Mismatches 74; Indels 0; Gaps 0;  
QY 35 GCACCTGCTGACGATCTTCGCGCGGTGACGAGTGTGCTGAGGCGCTCGGCGCCGACG 94  
DB 466 GAACCTGATCCACGATCTCACCGAATTGAAGTTACGCTCCAAACCGCCTGACCAAGTTG 407  
QY 95 CGCAGAACTGGACAGCGCAAGCGCGACCCCGCTGAGACGAGACGCGGTGTAGCGGT 154  
DB 406 CACGAATGGGACAGCGCATGCGGACCAAGCTTGTGCAAGAAATACAGCGCTGAGCGAT 347  
QY 155 CGTGCGACGAGATCGGACCGCTCTCGATGAATTGCGGTGACCAAGCTCGAAGCAAGAACT 214  
DB 346 CTGAATGCGGATGAGAAAGAGATCAAGCAAAAGTGCATATGAGAGCTTCATTAACGCT 287  
QY 215 T 215  
DB 286 T 286

RESULT 15  
ACA00654/c  
ID ACA00654 standard; DNA; 1098 BP.

XX ACA00654;  
AC  
XX 03-JUN-2003 (first entry)  
DT  
XX C. glutamicum derived ORF SEQ ID 645.  
DE  
XX

**KM** Coryneform; nucleic acid array; fermentation; culture; ds.  
**XY**

OS  
YY

*Corynebacterium glutamicum.*

PN DE10128510-A1.

PD 19-DEC-2002.

PF 13-JUN-2001; 2001DE-01028510.

PR 13-JUN-2001; 2001DE-01028510.

PA (DEGS ) DEGUSSA AG.

PI Farwick M, Moeckel B, Pfefferle W, Bathe B, Hutmacher K;

DR WPI; 2003-279970/28.

PT New nucleic acid array useful for monitoring mRNA expression of  
PT *Corynebacterium glutamicum* during fermentation, comprising nucleic acid  
PT from *Corynebacterium glutamicum*.  
xx

PS Claim 1; Page 269; 709pp; German.

CC This invention describes a novel nucleic acid array involving  
CC Corynebacterium glutamicum polynucleotides. The arrays are used to  
CC analyse C. glutamicum, particularly for monitoring a fermentation process  
CC to determine expression levels of C. glutamicum cellular mRNA. Such  
CC monitoring particularly differentiates between expression levels of  
CC different strains of C. glutamicum and allows the adjustment of different  
CC culture and fermentation conditions. ACA00010-ACA02188 represent C.  
CC glutamicum derived polynucleotides described in the disclosure of the  
CC invention

SQ Sequence 1098 BP; 221 A; 295 C; 304 G; 278 T; 0 U; 0 Other;

Query Match	10.5%	Score 62.6;	DB 8;	Length 1098;
Best Local Similarity	59.1%;	Pred. No. 3.1e-05;		
Matches 107; Conservative	0;	Mismatches 74;	Indels 0;	Gaps 0;

QY	35	GCACCTGCTGCAGATCTGACCCGGTAGGAGGATGCGCTTAGAGCCGTCGGGGCCGACG	94
Db	466	GAACCTGATCAAGATCTACCGATTGAAGGTTACGGTCCAAACCGCCTGACCAATTG	407
QY	95	CGCAGATCTGGCA CGCGAAGGCGCA CCCCCTCTGAGACGACGCGAGCGTGTAGCCGT	154
Db	406	CACAGATGGGCA CGGCATGCCGCAACAGCTTGCGAAGAAATACACAGCGTGGAGGAT	347
QY	155	CGTGGCGAGCGATCCGCACCGTCTCGATGAATTGCCGTGACCAAGCTCGAACGAACT	214
Db	346	CTGAAATAGCGCATGAGAAACAGACTCAAGCAAGTCCATCATGAGCTTCATAAACGCT	287
QY	215	T 215	
Db	286	T 286	

Search completed: April 7, 2006, 03:16:51  
Job time : 343.371 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:24:14 ; Search time 2885.48 Seconds

(Without alignments)  
11760.789 Million cell updates/sec

Title: US-10-611-442-2\_COPY\_17000\_17596

Perfect score: 597

Sequence: 1 aacacgagcctgtgacgc.....cacgaacgcgcagcgggcgc 597

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenBml:\*

1: gb\_ba:\*

2: gb\_in:\*

3: gb\_env:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pr:\*

9: gb\_ro:\*

10: gb\_scs:\*

11: gb\_sy:\*

12: gb\_un:\*

13: gb\_vl:\*

14: gb\_hlg:\*

15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	87.4	14.6	110000	1	BA000043_11
C 2	87	14.6	301332	1	AE017237
C 3	84.4	14.1	110000	1	BA000045_22
C 4	83.2	13.9	308050	1	SC0939124
C 5	81.2	13.6	207050	1	AL646063
C 6	78.4	13.1	110000	1	AP006618_42
C 7	77.8	13.0	110000	1	AE000516_31
C 8	77.8	13.0	110000	1	AE017333_17
C 9	77.8	13.0	110000	1	CP000002_17
C 10	77.8	13.0	318050	1	BX248344
C 11	77.6	13.0	348676	1	BX842581
C 12	77.6	13.0	1080	6	AR628287
C 13	77.6	13.0	13751	6	AR619704
C 14	77.2	12.9	110000	1	BA000030_32
C 15	76	12.7	10719	1	AE001946
C 16	75	12.6	6713	6	AR621649
C 17	75	12.6	6713	6	AR619398
C 18	75	12.6	110000	1	BX571965_17

19	74.6	12.5	347894	1	BX640431	BX640431 Bordercell
20	74.6	12.5	348642	1	BX640446	BX640446 Bordercell
21	74.6	12.5	348934	1	BX640417	BX640417 Bordercell
C 22	73.8	12.4	110000	1	AE014295_04	Continuation (5 of AX492787 Sequence AX553954 Sequence)
C 23	73.8	12.4	349980	6	AX492787	Continuation (14 o CP000010_13
C 24	73.4	12.3	110000	1	CP000010_14	Continuation (15 o CP000010_13
C 25	73.4	12.3	110000	1	CP000010_14	Continuation (21 o CP000010_13
C 26	73.4	12.3	110000	1	CP000010_14	Continuation (22 o CP000010_13
C 27	72.6	12.2	110000	1	CP000030_22	Continuation (23 o CP000030_22
C 28	72.2	12.1	110000	1	BA000035_20	Continuation (21 o CP000030_22
C 29	72	12.1	110000	1	CP000088_07	Continuation (8 of CP000088_07
C 30	69.6	11.7	110000	1	AE017282_30	Continuation (31 o AE016922 Chromobac
C 31	68	11.4	300217	1	AE016922	Continuation (15 o AP006840_14
C 32	67	11.2	110000	1	AP006840_14	Continuation (15 o CR555306_07
C 33	66.2	11.1	110000	1	CR555306_07	Continuation (25 o AP008231_24
C 34	65.8	11.0	110000	1	AP008231_24	Continuation (12 o AE017180_11
C 35	65.6	11.0	110000	1	CR931997_13	Continuation (14 o CR931997_13
C 36	65	10.9	110000	1	CP000089_32	Continuation (33 o PSEB1P
C 37	65	10.9	110000	1	AE004798	Continuation (19 o BX569690
C 38	64.8	10.9	17148	1	AE004798	Continuation (19 o BX248358
C 39	64.8	10.9	17148	1	AE004798	Continuation (19 o BX248358
C 40	64.2	10.8	349652	1	BX569690	Continuation (19 o BX248358
C 41	63.6	10.7	49742	1	AP008226_18	Continuation (19 o BX248358
C 42	63.2	10.6	348408	1	BX248358	Continuation (19 o BX248358
C 43	62.6	10.5	1098	6	BD164415	Continuation (22 o AX122298
C 44	62.6	10.5	1098	6	AX122298	Continuation (22 o AX122298
C 45	62.6	10.5	110000	1	BA000036_21	Continuation (22 o

## ALIGNMENTS

RESULT 1  
BA000043\_11/c  
WPCOMENT

Sequence split into 36 fragments LOCUS BA000043 Accession BA000043

Fragment Name	Begin	End
BA000043_00	1	110000
BA000043_01	100001	210000
BA000043_02	200001	310000
BA000043_03	300001	410000
BA000043_04	400001	510000
BA000043_05	500001	610000
BA000043_06	600001	710000
BA000043_07	700001	810000
BA000043_08	800001	910000
BA000043_09	900001	1010000
BA000043_10	1000001	1110000
BA000043_11	1100001	1210000
BA000043_12	1200001	1310000
BA000043_13	1300001	1410000
BA000043_14	1400001	1510000
BA000043_15	1500001	1610000
BA000043_16	1600001	1710000
BA000043_17	1700001	1810000
BA000043_18	1800001	1910000
BA000043_19	1900001	2010000
BA000043_20	2000001	2110000
BA000043_21	2100001	2210000
BA000043_22	2200001	2310000
BA000043_23	2300001	2410000
BA000043_24	2400001	2510000
BA000043_25	2500001	2610000
BA000043_26	2600001	2710000
BA000043_27	2700001	2810000
BA000043_28	2800001	2910000
BA000043_29	2900001	3010000
BA000043_30	3000001	3110000
BA000043_31	3100001	3210000
BA000043_32	3200001	3310000
BA000043_33	3300001	3410000
BA000043_34	3400001	3510000
BA000043_35	3500001	3544776



```
gene
CDS
    complement(6994..7788)
    /locus_tag="MAP2703c"
    complement(6994..7788)
    /locus_tag="MAP2703c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA05020.1"
    /db_xref="GI:41397160"
    /translation="MEVEIIPRLKEPLYRIYEMRLROELGAADLPRIAVICDGN
    RRMADAGDDVSYGRMGAVKIAEMLRMOAGIEMTVYLLISTENLQDPDELGL
    IRIITDVYEICAPANRMSVRYTGDELLIGESEPRRLRGAVESTPYEASHVNAVGY
    GGRREIVDAVRSLISKELANGAETAEELIADVADASEMLYTSGGDDPDLVIRTSGEQ
    RLSGFLMQSAVSEMMFTSAHWPAPFRVDFLRALRDYSQRHRYGK"
    7961..8734
    /locus_tag="MAP2704"
    7961..8734
    /locus_tag="MAP2704"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA05021.1"
    /db_xref="GI:41397161"
    /translation="MSSQSTIANPEPDHEALVPGNTVEQLPBGAAQVIGKPMRGMI
    HTFSAMLAIVTGATLVSVSMAASPAGSSTLIYAATAVLPVSAITYRVRKMTAA
    RTRMRKLDHSMIFVPIAGSYTPPARLVMTQNTGVVLCTVWGALAGALKVCMPTAP
    RMIGVPLVLLGLVAIWVAPTIIHQGVANAMVAVGVLYSVGGIFVALRMDDPMS
    TFGYHEFFALPAIALIHYIYAMWFAVFTGDHMLATP"
    complement(8744..9127)
    /locus_tag="MAP2705c"
    complement(8744..9127)
    /locus_tag="MAP2705c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA05022.1"
    /db_xref="GI:41397162"
    /translation="MPNDAKTDAIRNTHRYIELVAKGADDLVELYADDAVEDEVG
    GGVHIGROAIGHFYSALDNIQRECELLITLRVAGNEAAFFRLITVAGBRHMYLEPIDV
    WVDGNGKVAAMKAYVSPDDVSRP"
    complement(9149..11164)
    /locus_tag="MAP2706c"
    complement(9149..11164)
    /locus_tag="MAP2706c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA05023.1"
    /db_xref="GI:41397163"
    /translation="MSPADPSATNTLGRTSPYLRQHADNPVHMOQWTPQALADAAR
    DVPILISGYACMCHVMAHSEFDDPVAAAMNSGFVCKVDREERPIDIVYNNAT
    VALTGHGWPMTCLTPDGRPPFCGGYVKAFLDLSVATWRRBGEVEASRI
    AGELRMTSGLPAGADLATALCDENVAIILDRDVRHGGFGAGPFPSPALIELLR
    STERTGSAVTLQAVATGAMARGGITDQGGGFAYSDNMAVVFERKMYDNLIL
    LRLVAMWARTCDPLAARVATARFLDLDLGDGMSFLDADADGREGSYVWTPA
    QLTREVIGPDDGMAAVFGVTEGTGEHGSVLRPADPPTPRFDRVRAALLAARG
    RIOPGRDDKVTSMNGIATTAETASVALDDGDLADARCARSVLDLHVQGRBRA
    SLGGVVDGSAATLEBHAMLTGTLALYOLSAEDSWTATDILDTRLAHFADROGR
    WPTTADDAERLRLRPADPIDGATPSGASVTEALITRAHUVGERERIRLRAATTDLR
    AAVALTERAPRTAGHMLAVAEANVRGFLDIAVACDPARALADARLAPGALIVGG
    PDDSSBLLAGRDRVAGADAAYVCRGVCDLPVTRTELAALGABSP"
    complement(11151..11471)
    /locus_tag="MAP2707c"
    complement(11151..11471)
    /locus_tag="MAP2707c"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AA05024.1"
    gene
CDS
```

```

    /db_xref="GI:41397164"
    /translation="NMHLVLSGAVADGAPHHTGPDPRGRASPVGLIIVYLLISTLEL
    IASMRQLAKVESPDQHPERPDQAADBEITDAVDADADADGPDHANGSPRPBGSD
    DRPG"
    complement(11468..11889)
    /locus_tag="MAP2708c"
    gene
Query Match      14.6%; Score 87; DB 1; Length 301332;
Best Local Similarity 63.8%; Pred. No. 2.8e-05;
Matches 132; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
    gene
    complement(11468..11889)
    /locus_tag="MAP2708c"
    20 CGGGCCGAGCGCGACACCTGCTGCACATCGCCCGGTAGCAGTGTGCTTGAGGC 79
    Db 245970 CGCGCGCGCGCGCGCGACCTGCTGCAGATCTCGGCCCTTCACAGCTTGCGGCTCAGCC 246029
    Qy 80 CGTGGCGGCCGACGCGCAAACTGGACGCGAAGCGCAACCCGCTTAGAGCAGACGC 139
    Db 246030 CGCCCTGCGCGGTGCGCGCAAGACGGGCGCCATGCGCAGCCGCGCTGGAAGAATGC 246089
    Qy 140 AGCGCGGTGACCGTCTGTCGCGACGATCCGACCGCTTCGATGAATTGCCCTCGACCA 199
    Db 246090 ACACCGTGTGCGGTGCGGTAGCGCATGACGATCCGATCCAGTACCCATCCAGGG 246149
    Qy 200 GCTCGAACAAGAACTTGTGCTGTGCG 226
    Db 246150 CCGCGCAACAAGCTTTCGGGTCTGCG 246176
RESULT 3
BA000045_22
WPCOMMENT
Sequence split into 47 fragments LOCUS BA000045 Accession BA000045
Fragment Name Begin End
BA000045_00 1 110000
BA000045_01 10001 210000
BA000045_02 20001 310000
BA000045_03 30001 410000
BA000045_04 40001 510000
BA000045_05 50001 610000
BA000045_06 60001 710000
BA000045_07 70001 810000
BA000045_08 80001 910000
BA000045_09 90001 1010000
BA000045_10 100001 1110000
BA000045_11 110001 1210000
BA000045_12 120001 1310000
BA000045_13 130001 1410000
BA000045_14 140001 1510000
BA000045_15 150001 1610000
BA000045_16 160001 1710000
BA000045_17 170001 1810000
BA000045_18 180001 1910000
BA000045_19 190001 2010000
BA000045_20 200001 2110000
BA000045_21 210001 2210000
BA000045_22 220001 2310000
BA000045_23 230001 2410000
BA000045_24 240001 2510000
BA000045_25 250001 2610000
BA000045_26 260001 2710000
BA000045_27 270001 2810000
BA000045_28 280001 2910000
BA000045_29 290001 3010000
BA000045_30 300001 3110000
BA000045_31 310001 3210000
BA000045_32 320001 3310000
BA000045_33 330001 3410000
BA000045_34 340001 3510000
BA000045_35 350001 3610000
BA000045_36 360001 3710000
BA000045_37 370001 3810000
BA000045_38 380001 3910000
BA000045_39 390001 4010000
```



gene  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
250..2490  
/gene="SC05495"  
/note="synonym: SC8D9.07"  
250..2490  
/gene="SC05495"  
/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa, similar to many eg. TR:087376 (EMBL:AF052518) g-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300aa containing several potential membrane spanning hydrophobic regions and the remainder of the protein showing similarity to the phosphodiesterase mentioned above. Also similar to a nearby gene on the same contid (SC8D9.22) fasta scores; opt: 736, z-score: 625.6, E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains TTA leu codon, possible target for hda regulation. Also contains Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46 and Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative phosphodiesterase"  
/protein\_id="CAB37571.1"  
/db\_xref="GI:4467246"  
/db\_xref="GOA:Q92584"  
/db\_xref="InterPro:IPR000160"  
/db\_xref="InterPro:IPR00163"  
/db\_xref="UniProt/TREMBL:Q92584"  
/translat="MKEPTESAAGSRLLRLRMADAMHVSRLGQDGVSSMTGTRVY  
POAVRPHGHGQHPHGLVNDPGRSLHSWPAIPAAVATAFVLAGGYRAFSQGHLL  
PFAGVNGLSLAVTIGIVVGHVLMGRSNWGGTSGAALTAVLLVGMIPAGVSLIGI  
VAVLGLIRRRGRGQVLRGAVIDLGIAGAVYVSGSVSVSRPMDPMTGVYARAR  
KVLVAVALVAVNTRLLVITQTPRGLTARLAKTVLRGVAVALLGIAPITCVYAAVA  
KVLVAVALVAVNTRLLVITQTPRGLTARLAKTVLRGVAVALLGIAPITCVYAAVA  
ARAALMLIDLRFRSVNDTGLHAGDRLILCTQARLRALPGAEAAALGQDEFAVLL  
PVADTSATATLARSVLALSSPMIDDLTVTLRKSAGVAAPPHALDAEGLLRADVLL  
AMYQAKRRRTGVYVESKRSNTDPRLGLDLDLRALDAHVOVLYQKVFEDQVAGV  
LEALTRWYHPRGKYPDPERFAIASSSSIMHLYETVLETLGVARARAGLLPVPV  
VAVSRDVTHTGFAGSVARLARHGVAPGAQLQLETTETVALLIEDPRADTINALTGRK  
VMSLIDDTGTSYSLVHLRRLPVSESLKIDRSFVARLALDNDDAELVETVLVASLGI  
LVAVSGVDDGTETWERTLRDLCGDAAVQGMVLAAMPPEETAWLLARGSGWRAAALL  
AASAPR"

misc\_feature  
1141..1635  
/gene="SC05495"  
/note="Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46"  
1420..1422  
/gene="SC05495"  
/note="TTA leu codon, possible target for hda regulation"  
1672..2406  
/gene="SC05495"  
/note="Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83"  
2774..3145  
/gene="SC05496"  
/note="synonym: SC8D9.08"  
2774..3145  
/gene="SC05496"  
/note="SC8D9.08, small hydrophobic membrane protein, len: 123 aa; no significant database similarities. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative small membrane hydrophobic protein"  
/protein\_id="CAB37572.1"  
/db\_xref="GI:4467247"

gene  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
250..2490  
/gene="SC05495"  
/note="synonym: SC8D9.07"  
250..2490  
/gene="SC05495"  
/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa, similar to many eg. TR:087376 (EMBL:AF052518) g-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300aa containing several potential membrane spanning hydrophobic regions and the remainder of the protein showing similarity to the phosphodiesterase mentioned above. Also similar to a nearby gene on the same contid (SC8D9.22) fasta scores; opt: 736, z-score: 625.6, E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains TTA leu codon, possible target for hda regulation. Also contains Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46 and Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative phosphodiesterase"  
/protein\_id="CAB37571.1"  
/db\_xref="GI:4467246"

misc\_feature  
1141..1635  
/gene="SC05495"  
/note="Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46"  
1420..1422  
/gene="SC05495"  
/note="TTA leu codon, possible target for hda regulation"  
1672..2406  
/gene="SC05495"  
/note="Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83"  
2774..3145  
/gene="SC05496"  
/note="synonym: SC8D9.08"  
2774..3145  
/gene="SC05496"  
/note="SC8D9.08, small hydrophobic membrane protein, len: 123 aa; no significant database similarities. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative small membrane hydrophobic protein"  
/protein\_id="CAB37572.1"  
/db\_xref="GI:4467247"

gene  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
250..2490  
/gene="SC05495"  
/note="synonym: SC8D9.07"  
250..2490  
/gene="SC05495"  
/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa, similar to many eg. TR:087376 (EMBL:AF052518) g-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300aa containing several potential membrane spanning hydrophobic regions and the remainder of the protein showing similarity to the phosphodiesterase mentioned above. Also similar to a nearby gene on the same contid (SC8D9.22) fasta scores; opt: 736, z-score: 625.6, E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains TTA leu codon, possible target for hda regulation. Also contains Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46 and Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative phosphodiesterase"  
/protein\_id="CAB37571.1"  
/db\_xref="GI:4467246"

misc\_feature  
1141..1635  
/gene="SC05495"  
/note="Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46"  
1420..1422  
/gene="SC05495"  
/note="TTA leu codon, possible target for hda regulation"  
1672..2406  
/gene="SC05495"  
/note="Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83"  
2774..3145  
/gene="SC05496"  
/note="synonym: SC8D9.08"  
2774..3145  
/gene="SC05496"  
/note="SC8D9.08, small hydrophobic membrane protein, len: 123 aa; no significant database similarities. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative small membrane hydrophobic protein"  
/protein\_id="CAB37572.1"  
/db\_xref="GI:4467247"

gene  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
250..2490  
/gene="SC05495"  
/note="synonym: SC8D9.07"  
250..2490  
/gene="SC05495"  
/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa, similar to many eg. TR:087376 (EMBL:AF052518) g-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300aa containing several potential membrane spanning hydrophobic regions and the remainder of the protein showing similarity to the phosphodiesterase mentioned above. Also similar to a nearby gene on the same contid (SC8D9.22) fasta scores; opt: 736, z-score: 625.6, E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains TTA leu codon, possible target for hda regulation. Also contains Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46 and Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative phosphodiesterase"  
/protein\_id="CAB37571.1"  
/db\_xref="GI:4467246"

misc\_feature  
1141..1635  
/gene="SC05495"  
/note="Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46"  
1420..1422  
/gene="SC05495"  
/note="TTA leu codon, possible target for hda regulation"  
1672..2406  
/gene="SC05495"  
/note="Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83"  
2774..3145  
/gene="SC05496"  
/note="synonym: SC8D9.08"  
2774..3145  
/gene="SC05496"  
/note="SC8D9.08, small hydrophobic membrane protein, len: 123 aa; no significant database similarities. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/product="putative small membrane hydrophobic protein"  
/protein\_id="CAB37572.1"  
/db\_xref="GI:4467247"

gene  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
250..2490  
/gene="SC05495"  
/note="synonym: SC8D9.07"  
250..2490  
/gene="SC05495"  
/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa, similar to many eg. TR:087376 (EMBL:AF052518) g-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300

/db\_xref="UniProt/TREMBL:Q9Z583"  
/translation="MTVATYLAIVLTFCLFLGYRFLQFPASATGYPADAGDA  
AAYLVKGVGDTGLGVVGLALVLAAPGARPRAMFGLCVALVPLADTLVLHGGTKANA  
FGIHPATVAVLISAGLLFAV"  
3177.3671  
/gene="SC05497"  
/note="synonym: SC8D9.09"  
3177.3671  
/gene="SC05497"  
/note="SC8D9.09, unknown, len: 164aa; similar to many of  
undefined function eg. TR:006336 (EMBL:Z95390)  
hypothetical protein from Mycobacterium tuberculosis (177  
aa) fasta scores; opt: 171, z-score: 213.4, E(): 0.00014,  
(34.9% identity in 109 aa overlap)."  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CAB37573.1"  
/db\_xref="GI:4467248"  
/db\_xref="GOA:Q9Z582"  
/db\_xref="InterPro:IPR003313"  
/db\_xref="InterPro:IPR007113"  
/db\_xref="UniProt/TREMBL:Q9Z582"  
/translation="MSLVKPFDESVVVRKAEAVNGRAPTVRLADSSGTGALST  
QRYTLTAGADGAKPHMDNSAEVFLDGDADVLGDEVLTAGGDLIVPPKPHAF  
AAYGADADLLIVLAPGVRFYFRHLQRIALGATPESLLVQLVDNHLSSAAMD  
ARRG"  
3816.4112  
/gene="SC05498"  
/note="synonyms: galc, SC8D9.10"  
3816.4112  
/gene="SC05498"  
/note="SC8D9.10, probable Glu-tRNA-Gln amidotransferase  
subunit C, galc, len: 98aa; similar to many eg. TR:006492  
(EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit C,  
galc, from Bacillus subtilis (101 aa) fasta scores; opt:  
233, z-score: 294.8, E(): 4.1e-09, (42.1% identity in 95  
aa overlap)."  
/codon\_start=1  
/transl\_table=11  
/product="probable Glu-tRNA-Gln amidotransferase subunit C"  
/protein\_id="CAB37574.1"  
/db\_xref="GI:4467249"  
/db\_xref="GOA:Q9Z581"  
/db\_xref="InterPro:IPR003837"  
/db\_xref="InterPro:IPR004415"  
/db\_xref="UniProt/Swiss-Prot:Q9Z581"  
/translation="MPGITREVAHLARLARLKCBELEHFAQGLDIIICAVARYSE  
VADQVPTSHPLPLTNVMPDEVRPSLFPBQALSGAPAQEQRFKPIIGSE"  
4116.5609  
/gene="SC05499"  
/note="synonyms: gata, SC8D9.11"  
4116.5609  
/gene="SC05499"  
/note="SC8D9.11, probable Glu-tRNA-Gln amidotransferase  
subunit A, gata, len: 497aa; similar to many eg. TR:006491  
(EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit A,  
gata, from Bacillus subtilis (486 aa) fasta scores; opt:  
1443, z-score: 1512.5, E(): 0, (52.0% identity in 465 aa  
overlap)."  
/codon\_start=1  
/transl\_table=11  
/product="probable Glu-tRNA-Gln amidotransferase subunit"  
/protein\_id="CAB37575.1"  
/db\_xref="GI:4467250"  
/db\_xref="GOA:Q9Z580"  
/db\_xref="InterPro:IPR001120"  
/db\_xref="InterPro:IPR004412"  
/db\_xref="UniProt/Swiss-Prot:Q9Z580"  
/translation="MSDIKTLTAETAKIASELTAVOVTEAHLEIAVDEKTHAF  
LHVREGALQAQAVDEKREKGLGLGLVPLADITFTTBTGIPPTVSKLIEGIP  
PDVATYTKLKADVVILGKTNDDRFAMGSSSTENSAVGFQNPMDLTKIPGSGGSS  
AALAFQAPLAIQITGSGIRQPAVTVGTGVKPYGVSRYGVMAVFSSSLDGGRCGA

RVTLDAALHEVIAGHPMDSTSIDAPVPAVVEAARNGSVGNRGVVKOPRGEYO  
GVORDBSESEITLSKSGARIVELDCSPDLASAYLIASBGSNINARIDGARVGR  
VGDGTHSEETLSLRAGREKVRIRIMIGYALSSGYDAYGAQYRTIKRE  
FERAPROVDIVPTPTTTFATIGERADDMANTLADCTIPNLGAANMSLPCGLA  
PEDNLPVGLQIILAPAMKDRLYKVAVAEAPVEKWHPLIBRPSL"  
5606.5845  
/gene="SC05500"  
/note="synonym: SC8D9.12"  
5606.5845  
/gene="SC05500"  
/note="SC8D9.12, putative membrane protein, len: 79 aa.  
Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="CAB37576.1"  
/db\_xref="GI:4467251"  
  
Query Match 13.9%; Score 83.2; DB 1; Length 308050;  
Best Local Similarity 62.5%; Pred. No. 0.00012;  
Matches 130; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
  
QY 14 TGAAGCCGCGGCCGAGCGCCAGCACTGTCTGACGATCTTGGCCCGTAGCAGTGTGCT 73  
Db 162508 TGAAGCCGCGGCATGCGCGTCCACGATCTGTGTGACGATCTTGGCGGTGACAGGTGTGCGGT 162449  
QY 74 TGAAGCCGCTGCGCGCCGACGCGCAGAACTGTGCAACGCGAAGCGCACCCCGCTGAGACG 133  
Db 162448 CCAAGCGCGCGCTGCGCGGTGCGCCAGAAACGCGAGTTCATCCCGCAGCCCGCTGAGACG 162389  
QY 134 AGACGAGCGCGGTGATGCGGTGCGTGGCGACGATCCGCAACCGCTCGATGAATGTCGCT 193  
Db 162388 TGAATGACATGATGACACCGCGTCCGGGTAGGCAATGAGACCGACCTGACGAGCGTCCGT 162329  
QY 194 CGACGAGCTCGAACAGAACTTGTGCT 221  
Db 162328 CGAAGCGCTTCCACAGCGTCTTCCGCGT 162301  
  
RESULT 5  
AL646063/LOCUS 207050 bp DNA linear BCT 16-APR-2005  
DEFINITION Ralstonia solanacearum GMI1000 chromosome, complete sequence;  
ACCESSION AL646063 AL646052  
VERSION AL646063.1 GI:17428161  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE  
AUTHORS 1  
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangent,S.,  
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Catolico,L.,  
Chandler,M., Choise,N., Claudel-Renard,C., Cunac,S., Demange,N.,  
Gaspier,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,  
Siglier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,  
Weissenbach,J. and Boucher,C.A.  
Genome sequence of the plant pathogen Ralstonia solanacearum  
Nature 415 (6871), 497-502 (2002)  
PUBMED 11823852  
TITLE 2 (bases 1 to 207050)  
REFERENCE Boucher,C.A.  
AUTHORS Direct Submision  
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, UMG CNRS  
118 Route de Nardonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire

COMMENT INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Bouchere@coulouse.inra.fr  
http://sequence.coulouse.inra.fr/R.solanacearum.html.

FEATURES  
Source  
1. 207050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GMI1000"  
/db\_xref="taxon:305"  
complement(332..1561)  
/gene="RSC1148"  
/note="synonym: RS04604"  
complement(332..1561)  
/gene="RSC1148"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD14850.1"  
/db\_xref="GI:17428162"  
/db\_xref="InterPro:IPR003347"  
/db\_xref="InterPro:IPR007113"  
/db\_xref="UniProt/TREMBL:Q8Y096"  
/translation="MNELPLVAAAGVPGTAAAGPAQLGSLSPDFMRTWKKPL  
LIRQALSPEDMRALIFPLSPALIKLARREVSRLIQTRGRWTFNNGPNERPLS  
RKARWTLTVGVNLAPEPAVALMORFPIPARLDVMIISFATDGGVGHPSYDV  
PLAHGRRRRIRISNDLTIVPLPLILNFPESEFLLPGDMLTLPHYAHGV  
ASGDMTYSGVGRSPSYRELAGHFGFISQTLNPNPDREGYTPDEQCPVRRPELPT  
AMVRALQKLNALNTPRLVGSFSAVLTSPDHVEFTQPLSLARTAPARKEGIV  
LDARTQALYDAQRFWINGDTPEPSTLLAMLSLSDRGASAAVDAADLPDLMDTL  
HAWYEGWLQLAGPGAH"  
1573..2097  
/gene="g1yD"  
/note="synonym: RS04602, RSC1149"  
1573..2097  
/gene="g1yD"  
/BC\_number="5.2.1.8"  
/function="cell processes; chaperoning"  
/note="Product confidence : probable  
Gene name confidence : probable  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS  
ISOMERASE (ROTAMASE) PROTEIN"  
/protein\_id="CAD14851.1"  
/db\_xref="GI:17428163"  
/db\_xref="GOA:Q8Y095"  
/db\_xref="InterPro:IPR001179"  
/db\_xref="UniProt/TREMBL:Q8Y095"  
/translation="MKIANTVSAVAYKLSAQGVIESDSDEPMVYLHGVDGTEPKI  
EHALDGHGSEFTKQLBPDPAFGEYDQLIKIEPRDRPPELLEVMGFEGVPEGDD  
BDALITVTVDAVEDKVVLDGNHPLAGMALRFLRATVDRAPAEIEIHRGHVANGGLE  
VADDEDERPPRTIH"  
complement(2152..3216)  
/gene="RSC1150"  
/note="synonym: RS04601"  
complement(2152..3216)  
/gene="RSC1150"  
/function="miscellaneous; unknown"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"

/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE TRANSMEMBRANE PROTEIN"  
/protein\_id="CAD14852.1"  
/db\_xref="GI:17428164"  
/db\_xref="UniProt/TREMBL:Q8Y094"  
/translation="MARPGAFAFRMRGIALAALMLATAAPAPKHAATAPAAPACE  
TFAVGVNVPQRPEDVAPARALDAIDHEPAFVHLGNLGRDESCTDNILSEKHDL  
DSITPLIYRGDHDSDCGRPTAGRPDIERLLRLADLFYPDNALGQRTVMRQS  
DQKFRSYRNNARMTGGVLFYTLNVGDNNTKTGRRGXEYDELRAROMYARAF  
ALNRKLPQGVITMQADPLFDEGERRRPNLLDGLRRRPHDGYALRQGLARAR  
EYDAVLLIHGASQSFILDRPLKDESGRPDHPWARTFQSPITLDQMLEYVTPGTAP  
LFHIRARLASDIAPSTPQQ"  
complement(3216..5864)  
/gene="muts"  
/note="synonym: RS04600, RSC1151"  
complement(3216..5864)  
/gene="muts"  
/function="macromolecule metabolism; macromolecule  
synthesis, modification; dna - replication, repair,  
restr./modif."  
/note="Product confidence : probable  
Gene name confidence : probable  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE DNA MISMATCH REPAIR PROTEIN"  
/protein\_id="CAD14853.1"  
/db\_xref="GI:17428165"  
/db\_xref="GOA:Q8Y093"  
/db\_xref="InterPro:IPR000432"  
/db\_xref="InterPro:IPR005748"  
/db\_xref="InterPro:IPR007695"  
/db\_xref="InterPro:IPR007696"  
/db\_xref="InterPro:IPR007860"  
/db\_xref="InterPro:IPR007861"  
/db\_xref="UniProt/Swiss-Prot:Q8Y093"  
/translation="WASHSAPALSIDKIGFDFKHTPMQOYLKSNHPTTVPYRM  
GDFEYFPEDAEEKARLIDITLTSRGTNGQPIWAGVPEFAVQYLAKLVKDESVA  
IGBOIGDPATTKGVKERRVAVVPTGLTAAALSDVNNHLLAIQIPGRGAAPLV  
GLANTLVGSELRLMECSABOLDRELRIRAEVADATLNAVAFVARRLPDMVH  
DVEGARIRIEOLGASVAVGCELTALAAAGALANTYATGOSIRHYIGLVESH  
ESRIGDPTARRRLLETITLROGESPTLSLDTCAATSGSRLLRHTLHPLRDRAV  
PQARQQAIEVLADGWSLKTLLTSDVERITRRALLSRRPDSISRTTLARPE  
VRDQLPQSDAQGLDLAALAIPEDAHALQRAVMAEPMAVVDGGVIAAGVADLDL  
ELADISENCQFLVDLBARERERTGIANLREYVRVGFYIEVNGQAAKVPDDYRR  
QTLKNAERYITPELKAFFEDKALSQDRLSREKYLEDLQKLPHLFVLKRIAAALA  
ADVIATTLAERAHALSWSRPALTDAPGIELRABHPVYBQVQEVANDCYLQSTRKL  
LITVPMNGKSTFMRQTAIVYLAAYGAPVAPAAATGPDRIETRIGADDLAAGR  
STFVETNEAAAILHRAIPNSLVIMDEIGRGTSTFDGIALAMALARHLSNRSHITLF  
ATHYFELTQLPQERAAQAAVHLSAVEHQDGIYFLHVAQEGPASQSYGIVQVQLGVPO  
PVIRARAKRLAMEQHSADTGATPOLDLFALPSPSDDAEAAAPAPSAALALDG  
IDPSPMPRDAIDALYRLKALDSASA"  
complement(5960..6853)  
/gene="RSC1152"  
/note="synonym: RS04599"  
complement(5960..6853)  
/gene="RSC1152"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD14854.1"

```

/db_xref="GI:17428166"
/db_xref="GOA:O8Y092"
/db_xref="InterPro:IPR000847"
/db_xref="InterPro:IPR005119"
/db_xref="InterPro:IPR011991"
/db_xref="UniProt/TREMBL:O8Y092"
/translation="MKTTLEELAFRTVTVGTAAAEQIGQTGVSGBALRRLEEK
LETTLLSTTRRLATDEGINFLAHARGIILAAVEASEELLALRRQQPAGLRNANMP
FMHLVIADLVADFRAQFQPIELENTDRIITIDLEBTDVAIRIGLRSTLERARLG
TRLHLVIAVLPAKAGRPRTVDALRRHALLFGQPESLNHWPLRGHGGRIRTPTL
KASSETLRRLALOGAGIVCLADFTMRABDRAGDLVGVLFEALIVEIQCPHFANYENT
QLASITTCETLDVLEKRLAQGE"
6966 . 8168
/gene="RScl153"
/note="synonym: RS04598"
6966 . 8168
/gene="RScl153"
/function="cell processes; transport of small molecules"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
Predicted by Homology"
13.6%; Score 81.2; DB 1; Length 207050;
Best Local Similarity 61.2%; Pred. No. 0.00027;
Matches 131; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
```

Query Match	13.6%	Score 81.2	DB 1	Length 207050
Best Local Similarity	61.2%	Pred. No. 0.00027		
Matches 131	Conservative	0	Mismatches 83	Indels 0
			Gaps	0
OY	8	G C C T G T T A G C C C G G C C C G A G C G C G A C A C T T G C T G C A C G A T C T G C C C G G T A G C A G G T	67	
Db	77091	G C T T G C C C A T T G G C G A A C T C G G G C A T C A C A G A C T G G C C G A C G A T C T G C C T G T G A G G T	77032	
OY	68	G T G C T T T A G G G C G G T C G C C G G C C C G A C G C G G A A A C T T G C A C G G A A G G C G A C C C G C C T	127	
Db	77031	T G C C C G A G A A C C C T G C T T G C C G G T G C A A C A G A A C C G G C A G T T G A C G G C G A C C C G C C T	76972	
OY	128	G A G A C G A G A C G C A G C G C G T T A G C C G T C G T G G C G A C G A T C C G A C C G T T C G A T G A A A T	187	
Db	76971	G C G A A G A C A C G C A C A G G T G C C G G C G T C T C T C G G G A A T G T A C A C G T C T G A C G G C A T	76912	
OY	188	T G C C G T G A C C A G C T G C A A C A G A A C T T T G T C C T	221	
Db	76911	T G C C G A C C C A C G T G A C A G C A C A C T T G C G C C T	76878	

Sequence split into 61 fragments	LOCUS	AP006618	Accession	AP006618
Fragment Name	Begin	End		
AP006618_00	1	110000		
AP006618_01	100001	210000		
AP006618_02	200001	310000		
AP006618_03	300001	410000		
AP006618_04	400001	510000		
AP006618_05	500001	610000		
AP006618_06	600001	710000		
AP006618_07	700001	810000		
AP006618_08	800001	910000		
AP006618_09	900001	1010000		
AP006618_10	1000001	1110000		
AP006618_11	1100001	1210000		
AP006618_12	1200001	1310000		
AP006618_13	1300001	1410000		
AP006618_14	1400001	1510000		
AP006618_15	1500001	1610000		
AP006618_16	1600001	1710000		
AP006618_17	1700001	1810000		
AP006618_18	1800001	1910000		
AP006618_19	1900001	2010000		
AP006618_20	2000001	2110000		
AP006618_21	2100001	2210000		
AP006618_22	2200001	2310000		
AP006618_23	2300001	2410000		

AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000
AP006618_49	4900001	5010000
AP006618_50	5000001	5110000
AP006618_51	5100001	5210000
AP006618_52	5200001	5310000
AP006618_53	5300001	5410000
AP006618_54	5400001	5510000
AP006618_55	5500001	5610000
AP006618_56	5600001	5710000
AP006618_57	5700001	5810000
AP006618_58	5800001	5910000
AP006618_59	5900001	6010000
AP006618_60	6000001	6021225

Continuation (43 of 61) of AP006618 from base 4200001 (AP006618 Nocardia farcinica IFM)

Fragment Name	Begin	End	LOCUS	AE000516	Accession	AE000516
AE000516_00	1	110000				
AE000516_01	100001	210000				
AE000516_02	200001	310000				
AE000516_03	300001	410000				
AE000516_04	400001	510000				
AE000516_05	500001	610000				
AE000516_06	600001	710000				
AE000516_07	700001	810000				

AE000516\_08 800001 910000  
AE000516\_09 900001 1010000  
AE000516\_10 1000001 1100000  
AE000516\_11 1100001 1210000  
AE000516\_12 1200001 1310000  
AE000516\_13 1300001 1410000  
AE000516\_14 1400001 1510000  
AE000516\_15 1500001 1610000  
AE000516\_16 1600001 1710000  
AE000516\_17 1700001 1810000  
AE000516\_18 1800001 1910000  
AE000516\_19 1900001 2010000  
AE000516\_20 2000001 2110000  
AE000516\_21 2100001 2210000  
AE000516\_22 2200001 2310000  
AE000516\_23 2300001 2410000  
AE000516\_24 2400001 2510000  
AE000516\_25 2500001 2610000  
AE000516\_26 2600001 2710000  
AE000516\_27 2700001 2810000  
AE000516\_28 2800001 2910000  
AE000516\_29 2900001 3010000  
AE000516\_30 3000001 3110000  
AE000516\_31 3100001 3210000  
AE000516\_32 3200001 3310000  
AE000516\_33 3300001 3410000  
AE000516\_34 3400001 3510000  
AE000516\_35 3500001 3610000  
AE000516\_36 3600001 3710000  
AE000516\_37 3700001 3810000  
AE000516\_38 3800001 3910000  
AE000516\_39 3900001 4010000  
AE000516\_40 4000001 4110000  
AE000516\_41 4100001 4210000  
AE000516\_42 4200001 4310000  
AE000516\_43 4300001 4400001  
Continuation (32 of 44) of AE000516 from base 3100001 (AE000516 Mycobacterium tuberculosis)

Query Match 13.0%; Score 77.8; DB 1; Length 110000;  
Best Local Similarity 61.7%; Pred. No. 0.0011;  
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

OY 26 CGAGCGCAGCAGCTGTGCAAGATCTGCCCGGTAGAGGATGTGCTTGAGGCGCTGCG 85  
DB 8434 CGCCGCGCAGCAGCTGTGCAAGATCTGCCCGGTAGAGGATGTGCTTGAGGCGCTGCG 84423  
OY 86 GAGCGCGCAGCAGCTGTGCAAGATCTGCCCGGTAGAGGATGTGCTTGAGGCGCTGCG 145  
DB 84424 GAGCGCGCAGCAGCTGTGCAAGATCTGCCCGGTAGAGGATGTGCTTGAGGCGCTGCG 84483  
OY 146 TGTAGCGCTGTGCGGAGGATCTGCCCGGTAGAGGATGTGCTTGAGGCGCTGCG 205  
DB 84484 TGTAGCGCTGTGCGGAGGATCTGCCCGGTAGAGGATGTGCTTGAGGCGCTGCG 84543  
OY 206 ACAGGACCTTGTGCTGCG 226  
DB 84544 ACAAGCTTTTCAAGCTTGGC 84564

RESULT 8  
AE017333\_17/c

WPCOMMENT

Sequence split into 43 fragments LOCUS AE017333 Accession AE017333

Fragment Name	Begin	End
AE017333_00	1	110000
AE017333_01	100001	210000
AE017333_02	200001	310000
AE017333_03	300001	410000
AE017333_04	400001	510000
AE017333_05	500001	610000
AE017333_06	600001	710000
AE017333_07	700001	810000
AE017333_08	800001	910000

AE017333\_09 900001 1010000  
AE017333\_10 1000001 1110000  
AE017333\_11 1100001 1210000  
AE017333\_12 1200001 1310000  
AE017333\_13 1300001 1410000  
AE017333\_14 1400001 1510000  
AE017333\_15 1500001 1610000  
AE017333\_16 1600001 1710000  
AE017333\_17 1700001 1810000  
AE017333\_18 1800001 1910000  
AE017333\_19 1900001 2010000  
AE017333\_20 2000001 2110000  
AE017333\_21 2100001 2210000  
AE017333\_22 2200001 2310000  
AE017333\_23 2300001 2410000  
AE017333\_24 2400001 2510000  
AE017333\_25 2500001 2610000  
AE017333\_26 2600001 2710000  
AE017333\_27 2700001 2810000  
AE017333\_28 2800001 2910000  
AE017333\_29 2900001 3010000  
AE017333\_30 3000001 3110000  
AE017333\_31 3100001 3210000  
AE017333\_32 3200001 3310000  
AE017333\_33 3300001 3410000  
AE017333\_34 3400001 3510000  
AE017333\_35 3500001 3610000  
AE017333\_36 3600001 3710000  
AE017333\_37 3700001 3810000  
AE017333\_38 3800001 3910000  
AE017333\_39 3900001 4010000  
AE017333\_40 4000001 4110000  
AE017333\_41 4100001 4210000  
AE017333\_42 4200001 422645  
Continuation (18 of 43) of AE017333 from base 1700001 (AE017333 Bacillus licheniformis)

Query Match 13.0%; Score 77.8; DB 1; Length 110000;  
Best Local Similarity 63.8%; Pred. No. 0.0011;  
Matches 116; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 37 ACCTGTGCAAGATCTGTGCGCGGTAGAGGATGTGCTTGAGGCGCTGCGCGCGGAGCG 96  
DB 55396 ACCTGTGCAAGATCTGTGCGCGGTAGAGGATGTGCTTGAGGCGCTGCGCGCGGAGCG 55337  
OY 97 CAGAGTGTGCAAGATCTGTGCGCGGTAGAGGATGTGCTTGAGGCGCTGCGCGCGGAGCG 156  
DB 55336 CAGAGTGTGCAAGATCTGTGCGCGGTAGAGGATGTGCTTGAGGCGCTGCGCGCGGAGCG 55277  
OY 157 TGGGAGGATCTGTGCGCGGTAGAGGATGTGCTTGAGGCGCTGCGCGCGGAGCG 216  
DB 55276 TGGGAGGATCTGTGCGCGGTAGAGGATGTGCTTGAGGCGCTGCGCGCGGAGCG 55217  
OY 217 GTGCT 221  
DB 55216 ATTTG 55212

RESULT 9  
CP000002\_17/c

WPCOMMENT

Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Hewinson, G.  
 The complete genome sequence of *Mycobacterium bovis*  
 Unpublished  
 2 (bases 1 to 318050)  
 Garnier, T.  
 Direct Submission  
 Submitted (24-MAR-2003) Garnier T., Unité de Génétique Moléculaire  
 Bactérienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
 15, France. e-mail: t.garnier@pasteur.fr Submitted on behalf of the  
 Mycobacterium Bovis Sequencing teams, TB Research Group, Veterinary  
 Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,  
 Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut  
 Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
 Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28  
 rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES  
 source

Location/Qualifiers  
 1. 318050  
 /organism="Mycobacterium bovis AF2122/97"  
 /mol\_type="genomic DNA"  
 /strain="AF2122/97"  
 /db\_xref="taxon:233413"  
 complement(99. 875)  
 /locus\_tag="Mb2882c"  
 complement(99. 875)  
 /locus\_tag="Mb2882c"  
 /EC\_number="1.-.-.-"  
 /note="Mb2882c, -, len: 258 aa. Equivalent to RV2857c,  
 len: 258 aa, from *Mycobacterium tuberculosis* strain H37Rv,  
 (100.0% identity in 258 aa overlap). Probable short-chain  
 dehydrogenase/reductase (EC 1.-.-.-), highly similar to  
 various dehydrogenases e.g. O88068|SC15.33c PROBABLE  
 DEHYDROGENASE (SDR FAMILY) from *Streptomyces coelicolor*

gene

```

/translation="MMMDLSQRLAGVAIVTGGSGGICLAGRRMRAGKATVGVGDV
EKGGAADSLSGLPVPTVYCDERDAVNGIPDCAAEVGLDGIADIANNGISIPPEENLIE
TELLAKQRYQDVNLKSVYLCRGAALRHMTLAGKGIYNTASTVAANGKSTQSISTAS
KGGVAANRHELEVFARQGIIRYNALCPGVNTPFLDQELFARPERARRMWHVPLGAP
ASPDLEIAAVALASDASFTASTFLVDGSSAYVPL"
complement(872..2239)
/gene="aIdc"
/locus_tag="Mb2883c"
complement(872..2239)
CDS
/gene="aIdc"
/locus_tag="Mb2883c"
/locus_tag="Mb2883c"
/BC_number="1.2.1.3"
/nc="Mb2883c.aIdc, len: 455 aa, Equivalent to Rv2858c,
len: 455 aa, from Mycobacterium tuberculosis strain H37Rv,
100.0% identity in 455 aa overlap). Probable aIdc,
aldehyde dehydrogenase (EC 1.2.1.3), similar to many e.g.
O88069|Scf13.34c PUTATIVE ALDEHYDE DEHYDROGENASE from
Streptomyces coelicolor (483 aa), FASTA scores: opt: 1872,
```



E(): 6.4e-109, (64.5% identity in 448 aa overlap);  
O9FAB1|ALPH|BT-ALDH ALDEHYDE DEHYDROGENASE from *Bacillus thermojovians* (497 aa), FASTA scores: opt: 1157, E(): 2.1e-64, (44.3% identity in 458 aa overlap); O33455|CMC P-CUMIC ALDEHYDE DEHYDROGENASE from *Pseudomonas putida* (494 aa), FASTA scores: opt: 1149, E(): 6.5e-64, (43.15% identity in 452 aa overlap);  
P40047|DHAS YEAS|ALD5|ALD5|YER073M ALDEHYDE DEHYDROGENASE from *Saccharomyces cerevisiae* (Baker's yeast) (519 aa), FASTA scores: opt: 1091, E(): 2.7e-60, (38.55% identity in 459 aa overlap);  
P80668|FEAB\_ECOLI|PADA|MA08|B1385 PHENYLACETALDEHYDE DEHYDROGENASE (EC 1.2.1.39) from *Escherichia coli* strain K12 (499 aa), FASTA scores: opt: 1074, E(): 3e-59, (42.2% identity in 462 aa overlap); etc. Also similar to many *M. tuberculosis* dehydrogenases e.g. P71823|Rv0768|MYC3363.13 (489 aa), FASTA score: (38.1% identity in 467 aa overlap).  
Contains P500687 Aldehyde dehydrogenases glutamic acid active site. BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY. "

/codon\_start=1  
/transl\_table=1  
/product="PROBABLE ALDEHYDE DEHYDROGENASE ALDC"  
/protein\_id="CAD96570.1"  
/db\_xref="GI:31619630"  
/db\_xref="GOA:O7TXP5"  
/db\_xref="InterPro:IPR020866"  
/db\_xref="UniProt/TrEMBL:O7TXP5"  
/translation="MSTTQLINPAIEEVLASVDHTANAVDAVQRAAQRKARIALA  
PACRAAGIRAAVAQALDELIALLEVANSCHPIVSAEMEGHVRDVLAFAPSPERL  
SGROI PVAGVDVTFENPMGVGIVTTPNPFMVLASWALPALAGNAVLAFAPSPERL  
LTMRIEGLAVBAGLDLLOVLPGKGVGERPVTHDIKIVFTSGTEKRVKVMG  
AAAYKVVTLELGGKSNIVFHDCLRAATAPAGVDNAGDCCASRLIVRSVY  
DRFELSLPAVHSTIVGDPGSRATKEMGLVSRARDKVAGVPPADAPARGTAPAR  
GFMPPTVLTPKRDRTVDEIFGPVVVVLTFDEADAIISLANDTAYGLSSIVTDLA  
SRLRVAAVSQNI SVNSHSVRFNTPFGGFKQSGVGRLEGPDAPLQFTKRVFLA  
VGEEM"

complement (2236, 3162)  
/locus\_tag="MD2884c"  
complement (2236, 3162)  
/locus\_tag="MD2884c"  
/EC\_number="2.-.-.-"  
/EC\_number="6.3.5.-"  
/note="MD2884c", len: 308 aa. Equivalent to Rv2859C, len: 308 aa, from *Mycobacterium tuberculosis* strain H37Rv, (100.0% identity in 308 aa overlap). Possible amidotransferase (EC 6.3.5.- or 2.-.-.-), equivalent (but longer 58 aa) to Q9CBU9|ML1573 POSSIBLE AMIDOTRANSFERASE from *Mycobacterium leprae* (249 aa), FASTA scores: opt: 1226, E(): 3e-64, (71.55% identity in 239 aa overlap). Also similar to other amidotransferases and hypothetical proteins, but shorter in N-terminus e.g. O88072|SC135.37 HYPOTHETICAL 25.3 KDA PROTEIN from *Streptomyces coelicolor* (242 aa), FASTA scores: opt: 683, E(): 1.2e-32, (47.65% identity in 235 aa overlap); AAK79730|Q97188|CAC1764 PREDICTED GLUTAMINE AMIDOTRANSFERASE from *Clostridium acetobutylicum* (241 aa), FASTA scores: opt: 458, E(): 1.6e-19, (32.95% identity in 246 aa overlap);  
AAK75201|Q970V9|SP1089 GLUTAMINE AMIDOTRANSFERASE CLASS I from *Streptococcus pneumoniae* (229 aa), FASTA scores: opt: 431, E(): 5.6e-18, (34.75% identity in 236 aa overlap); etc. Contains three 17 aa repeats at the N-terminus very similar to those in other *Mycobacterium tuberculosis* proteins e.g. Q10699|Y130\_MCTU|Rv2090|MT2151|MYC19.30 PUTATIVE 5'-3' EXONUCLEASE Rv2090 (EC 3.1.11.-). "

/codon\_start=1  
/transl\_table=1  
/product="POSSIBLE AMIDOTRANSFERASE"  
/protein\_id="CAD96571.1"  
/db\_xref="GI:31619631"  
/db\_xref="GOA:O7TXP4"  
/db\_xref="InterPro:IPR011697"

/db\_xref="UniProt/TrEMBL:O7TXP4"  
/translation="MDLSASRSQDPLRPASRLPSVSGDPLRPASRLPSV  
DGDPLRPASRLPSVGLTAYLEQVRTGWDIPAGLPADYFEGITMAG  
VAVLPQVPDPSVSGVDLSHALVITGGYDLDPAVQGPSPATPRPGDAMFE  
ALRGALQCMPTLGI CRGQVNVNVALGTHLHLPILIGSGHRAQNGVETRLPVHT  
ASTRIALDELIGESADPCTHQAQIDQVBSLVYASNDVGVIALRELPGTFVLAQVM  
HPRKSLDDLRLFPALVDASGVAGSQSAPR"  
complement (3143, 4516)  
/gene="glnA4"  
/locus\_tag="MD2885c"  
complement (3143, 4516)  
/gene="glnA4"  
/locus\_tag="MD2885c"  
/note="MD2885c", glnA4, len: 457 aa. Equivalent to Rv2860C, len: 457 aa, from *Mycobacterium tuberculosis* strain H37Rv, (100.0% identity in 457 aa overlap). Probable glnA4, glutamine synthetase class II (EC 6.3.1.2), similar to many glutamine synthetases e.g. O88070|SC135.35C from *Streptomyces coelicolor* (462 aa), FASTA scores: opt: 1947, E(): 8.2e-120, (64.15% identity in 452 aa overlap);  
O98H15|ML13074 from *Rhizobium loti* (*Mesorhizobium loti*) (465 aa), FASTA scores: opt: 1321, E(): 7.8e-79, (46.7% identity in 452 aa overlap); Q98EM0|ML1487 from *Rhizobium loti* (*Mesorhizobium loti*) (456 aa), FASTA scores: opt: 698, E(): 4.6e-38, (33.5% identity in 454 aa overlap);  
Q9CDL9|GLNA from *Lactococcus lactis* (subsp. *lactis*) (*Streptococcus lactis*) (446 aa), FASTA scores: opt: 633, E(): 8.2e-34, (32.45% identity in 456 aa overlap); etc.  
Also similar to three other potential glutamine synthetases in *Mycobacterium tuberculosis*:  
Q10378|GLN2 MYCTU|Rv2222c|MT2280|MYC190.33c|MYC1427 . 0.3c PROBABLE GLUTAMINE SYNTHETASE (446 aa), FASTA score: (31.1% identity in 453 aa overlap); Rv1878|glnA3 and Rv2220|glnA1. BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY. "

/codon\_start=1  
/transl\_table=1  
/product="PROBABLE GLUTAMINE SYNTHETASE GLNA4 (GLUTAMINE

Query Match 13.0%; Score 77.8; DB 1; Length 318050;  
Blast local similarity 61.7%; Pred. No. 0.00092;  
Matches 124; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 26 CGAGCGCCAGACCTGTCTGCAGATCTGCCCGGTAGCAGGTGTCTTGAGCCGTGCG 85  
DB 21725 CGCGGACGACGACCTGCTCGAGATCTCCGCGTGCAGATGAGTCAATCGGCCCT 21784  
QY 86 GCGCCGACGCGCAGATCTGCGACGCGGACGCGCCGCTTGAGACGACGCGGCGG 145  
DB 21785 GCGCAGTGGCACAGAACGCGGACGACGCGCGCGCTTCGAGAAATGCGACGCG 21844  
QY 146 TGTAGCCGCTGTGCGACGAGATCCGACGCTCTCGATGAATGCGCGTGCACGAGCTGCA 205  
DB 21845 TGTTCGCGCGGATAGGCGATTCGACGACGATTCGACATGCTGACGCGCCGCC 21904  
QY 206 ACAGGACTTGTGCTGTGCGC 226  
DB 21905 ACAACGCTTTGAGTGTGCGC 21925

RESULT 11  
BX842581 348676 bp DNA linear BCT 17-APR-2005  
LOCUS BX842581  
DEFINITION *Mycobacterium tuberculosis* H37Rv complete genome; segment 10/13.  
ACCESSION BX842581 AL008883 AL008967 AL021070 AL021309 AL123456  
Z74024 Z74697 Z81331 Z83018 Z83857 Z83858 Z83866 Z95150 Z95207  
VERSION BX842581.1 GI:41352756  
KEYWORDS complete genome.  
SOURCE *Mycobacterium tuberculosis* H37Rv  
ORGANISM *Mycobacterium tuberculosis* H37Rv  
Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
Corynebacteriineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
tuberculosis complex.







Db 5410 TTGTGCTGAAGCCGTGCTGTG 5387

RESULT 14  
BA000030\_32  
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragment Name	Begin	End
BA000030_00	1	110000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000
BA000030_45	4500001	4610000
BA000030_46	4600001	4710000
BA000030_47	4700001	4810000
BA000030_48	4800001	4910000
BA000030_49	4900001	5010000
BA000030_50	5000001	5110000
BA000030_51	5100001	5210000
BA000030_52	5200001	5310000
BA000030_53	5300001	5410000
BA000030_54	5400001	5510000
BA000030_55	5500001	5610000
BA000030_56	5600001	5710000
BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000

BA000030\_65 6500001 6610000  
BA000030\_66 6600001 6710000  
BA000030\_67 6700001 6810000  
BA000030\_68 6800001 6910000  
BA000030\_69 6900001 7010000  
BA000030\_70 7000001 7110000  
BA000030\_71 7100001 7210000  
BA000030\_72 7200001 7310000  
BA000030\_73 7300001 7410000  
BA000030\_74 7400001 7510000  
BA000030\_75 7500001 7610000  
BA000030\_76 7600001 7710000  
BA000030\_77 7700001 7810000  
BA000030\_78 7800001 7910000  
BA000030\_79 7900001 8010000  
BA000030\_80 8000001 8110000  
BA000030\_81 8100001 8210000  
BA000030\_82 8200001 8310000  
BA000030\_83 8300001 8410000  
BA000030\_84 8400001 8510000  
BA000030\_85 8500001 8610000  
BA000030\_86 8600001 8710000  
BA000030\_87 8700001 8810000  
BA000030\_88 8800001 8910000  
BA000030\_89 8900001 9010000  
BA000030\_90 9000001 9025608

Continuation (33 of 91) of BA000030 from base 3200001 (BA000030 Streptomyces avermitil.)

Query Match 12.9%; Score 77.2; DB 1; Length 110000;

Best Local Similarity 61.4%; Pred. No. 0.0014; Mismatches 78; Indels 0; Gaps 0;

Matches 124; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 14 TGAAGCCGCGCCGAGCGCAGCAGCTGTGCAAGATCTGCCGCTAGCAGGTGCT 73  
Db 21185 TGAAGCCGCGCCGAGCGCAGCTGTGCAAGATCTGTGCAAGATCTGCCGCTAGCAGGTGCT 21244

QY 74 TGAAGCCGCTGCGGCGCCGAGCGCAGAACTGCGCAGAGGCGACCCCGCTGAGACG 133  
Db 21245 CGAGGCCGCTGTGCGGCGCAGAGCGGCAAGTTCTACGACACCCGCTGAGACG 21304

QY 134 AGACGAGCGCGGTGAGCGCTGTGCGCAGCGATCCGACCCGCTGATGAAATTGCGCT 193  
Db 21305 TGAAGCAATCTGTACCCGCTGCGGCTAGCGATGAGCAGCAGTGAAGCGCTCCGT 21364

QY 194 CGACGAGCTGAAAGAACTT 215  
Db 21365 CGAAGCGCGCAGCGCTT 21386

RESULT 15  
AE001946 10719 bp DNA linear BCT 22-NOV-1999  
LOCUS  
DEFINITION Deinococcus radiodurans R1 section 83 of 229 of the complete  
chromosome 1.

ACCESSION AE001946 AE000513  
VERSION AE001946.1 GI:6458655  
KEYWORDS

SOURCE Deinococcus radiodurans R1  
ORGANISM Deinococcus radiodurans R1  
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;

REFERENCE 1 (bases 1 to 10719)

White, O., Bisen, J.A., Heidelberg, J.F., Hickey, B.K., Peterson, J.D.,  
Dodeson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
Moffatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
Vamathevan, J.J., Lam, P., McDonald, L., Uterback, T., Zaleski, C.,  
Makarov, K.S., Aravind, L., Daly, M.J., Fraser, C.M., et al.

Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans R1  
Science 286 (5444), 1571-1577 (1999)

TITLE JOURNAL  
PUBMED 10567266  
REFERENCE 2 (bases 1 to 10719)  
AUTHORS White, O., Bisen, J.A., Heidelberg, J.F., Hickey, B.K., Peterson, J.D.,



gene  
9316. .10020  
/gene="DR0939"  
CDS  
9316. .10020  
/gene="DR0939"  
/note="similar to GB:AL009126 percent identity: 51.93;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI0515.1"  
/db\_xref="GI:6458662"  
/translation="MSRVADIPATYGRVVTYIRVLELEAQNVLPAASSGELARRAGV  
TPQVRKDLITFRGFRIGITVAALRRRLRALGIDQTNVIVIGKRLGHALIANY  
PGASDYQFQNVGLFDVAQPDVGRERGLTIQHSQUGPPVAVSAGTPROVDMGLTVP  
ABHAQAAQALVAVAGVGLINFAPVVLQTDHLPRAPAAPGRBVTVENVDLAGMK  
RLAFVMLGPOAGPAEE"  
10017. .10643  
/gene="DR0940"  
10017. .10643  
/gene="DR0940"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAFI0517.1"  
/db\_xref="GI:6458664"  
/translation="MTMKQWFPVTLIGLLAGCTVAGSGRVNVGVVDGASQEV  
ATLTIPEKCDKXGCTPQKQELSLTDGQVTFPTAPRPSBAVTLEGYRVLSDRLDG

Query Match 12.7%; Score 76; DB 1; Length 10719;  
Best Local Similarity 60.8%; Pred. No. 0.0032;  
Matches 124; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy	17	CGCCGGGCGCGAGCGGACGCTGTCGACGATCGCCCGGTAGCAGGTGCGCTGA	76
Db	5385	CGCCCCCGGCGACCGGCGACGCTGCCCGCGATTCGCGGGGTGAGGTGCGCCGA	5326
Qy	77	GGCCGTGCGGCGCGCGACGCGACGCTGTCGACGAGGCGCACCCCGCTGAGACGAGA	136
Db	5325	AGCCCATCTTGCCTGCGGCGGCAAAAGCGGAGCGGCGCGACCCACCATGTCGAGA	5266
Qy	137	CGCAGCGGTGTAGCGCTGTCGCGACGATCCGACCGCTTCGATGAATTCGCTCGA	196
Db	5265	CGCAGTGTCTTGCCTGCGGTCGAGGTAGGATGTACACCGCTTCATTCGCCCGCTCGT	5206
Qy	197	CCAGCTCGAACAAGAACTTGTG	220
Db	5205	TGAGGCTGAACAGATTTTGACCG	5182

Search completed: April 7, 2006, 16:57:46  
Job time : 2893.48 secs

**This Page Blank (uspto)**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

OY	121	GAGGGCGGGCACTACCTGCGGACGACGAGGGCTGCGGGCCATGGAACCCGAGGGGCGATTC	180
Db	52720	GGTGTGGGCGACCTGGGCGCCGCGCGATCAGGTACCGGAGATGGCGCGGAGATGGGGGTC	52779
OY	181	GAGGAGCTGCGGACCAACCTCTGAGACGCGCGGGAGACCCGTGGGTGTGCTGTGTGACCTTGAC	240
Db	52780	GCGGCCATGGCAACGCGCGCTGTGAGAGAGACGACACCTGTGTACAGGTGGCGGAGATCGAC	52839
OY	241	CGGAGCGGATTGCTGTGAACTGTTCAACGCGCGCCGCGCGCGCCCTCTTTCGACGAACTTC	300
Db	52840	TGGGACCGGACCTACGTCGCGGTTCAACGCGACCCGCGCCGCTGCTGTGCTCGACCTTC	52899
OY	301	GGTGGGGTCCGCGCGCGG-----GCCAGAGAGACCGGT	333
Db	52900	CCCGAGGTGTGTGTGCTGTGTGACGCGCGGAGTCGAGCGAGAGAGCCCGCAGCGCGGAC	52959
OY	334	CAGGATTCGATCTCGCGCGGCGGTGCGATGCGCGGAGGCGGAAGTGCACGAGCAT	393
Db	52960	GAGCGCTGGGATTTGCGGAGCGGCTGCGCGGAGTGGCGGAGCCGACCGGAAACACGCG	53019
OY	394	GTCGCGCGGCTGTGCTCGAGCGAGGTGGCAGAGCGATGTGGGACCAACGCGACGCGT	453
Db	53020	TTTGCTGGAACCTGGTCCGGCGCAATGTGGCGGTGTACTGGAACACACCGGTCCGGAACGG	53079
OY	454	ATCGAGCGTAGCGTTCGCGCTTCCGTGACCTTGAGATTGCATCCAGCAACCGCGCTGCACTTG	513
Db	53080	ATCGACCCCGGCGGGGCTTTCACAGAGATCGGCTTGCATCGGTACCCGCGGTGGAATCG	53139
OY	514	CGGAACCGGCTTCGCGGGCGGTGACCGGGGATCCGGGTGCGCACACATCGTCTTTCACAC	573
Db	53140	CGCAACCGGCTCAACCGGGCCACCGGCGCTACGCGTCCCGCACCGTGAAGTTTCGACGAG	53199
OY	574	CCGAC	578
Db	53200	CCGAC	53204

```

US-11-056-470-1
/ RESULT 2
/ Sequence 1, Application US/11056470
/ Publication No. US20050261263A1
/ GENERAL INFORMATION:
/ APPLICANT: SANTI, Daniel V.
/ APPLICANT: MYLES, David C.
/ APPLICANT: TIAN, Zong-Qiang
/ APPLICANT: HUTCHINSON, C. Richard
/ APPLICANT: JOHNSON, Robert G., Jr.
/ APPLICANT: ZHOU, Yi-Qiang
/ APPLICANT: FENG, Li
/ TITLE OF INVENTION: BENZOQUINONE ANSAMYCINS
/ FILE REFERENCE: 300622007510
/ CURRENT APPLICATION NUMBER: US/11/056,470
/ PRIOR FILING DATE: 2005-02-11
/ PRIOR APPLICATION NUMBER: US 10/212,962
/ PRIOR FILING DATE: 2002-08-05
/ PRIOR APPLICATION NUMBER: US 60/310,779
/ PRIOR FILING DATE: 2001-08-06
/ PRIOR APPLICATION NUMBER: US 60/389,255
/ PRIOR FILING DATE: 2002-06-14
/ PRIOR APPLICATION NUMBER: US 60/393,929
/ PRIOR FILING DATE: 2002-07-03
/ PRIOR APPLICATION NUMBER: US 60/395,275
/ PRIOR FILING DATE: 2002-07-12
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 6210
/ TYPE: DNA
/ ORGANISM: Streptomyces geldanus
US-11-056-470-1

```

Query Match	25.0%	Score 150.4	DB 14	Length 6210
-------------	-------	-------------	-------	-------------

	Best Local Similarity	54.4%	Pred No. 4.7e-24	Matches 332	Conservative 0	Mismatches 266	Indels 12	Gaps 1
QY	1	GCTCTCAACGCGGCGGCAACGCTTCTTGACGCTTCCGCCGTGTGTGCGCGCAGT	60					
Db	5524	GCCAACTAACGCGCGCCGCAATGCTTCTCGAGCCTTCGCCAGATGGCGCACGCGCGC	5583					
QY	61	GGGCTGCGCGGTCACTCTGATTCGCTGTGGGTCTGTGGCGCGGCGAACAATGGCC-----	114					
Db	5584	CACCGGCGCCGCACTTCGTGTGGGTGTGGAGCGCAACACAGCAAGCTCACTCG	5643					
QY	115	-----GGTACCGAGGCGCGCGACTACTCTCGACGACGAGGAGCTGCGCGCCACTGACCGC	168					
Db	5544	ACGATGACCGCCACACCACTAACCGCGCACACGCGCGGGGGGTGCTGGGCAATGACAAC	5703					
QY	169	CAGCGGGCGATTCGAGAGCTGCGGACCACTTGAGCGCGGGGACCTGTGGGTGTGGTG	228					
Db	5704	GCCGAGGCGCTGCGGGGTGTTCGACACCGGGTTGCGGTCCGGGCGGCCCGCGCTGTGGCC	5763					
QY	229	GTGACCTTGACCGGGAGCGGTTCTGTGAACCTGTTCACCGCGCGCGCGCGCGCCCTC	288					
Db	5764	GCGAAGATTCAGCTCAACCGCCCTTGGCGCGCGCGGACGCCAGTTGTGCGCGTGTGGCG	5823					
QY	289	TTCCAGCAACTCGGTGGGTTCGCGCGCGGCGCGGAGAGACCGGTTCAGAACTGGATCTC	348					
Db	5824	GGAATGCGCGGTCCGGCGCGCGCGGACCGCGCGGACCGCGGCCCGCGCGGTGGCTG	5883					
QY	349	GCCCGCGCGCTGGCGGTGATGCGGGAGCGGAAAGTACAGGCAATGTGCGCGCGGTGTC	408					
Db	5884	TGCGGGGACGTGGCGCGGCTGTCCCGCGCGGCGAGGGAGTTCTCGCTCAACTGGGTG	5943					
QY	409	CGAGCCGAGGTGGCAGGGGTGCTGGGCGACGGGCAACGCGCGGAGCGGTATCGAGCGTGAAGTCA	468					
Db	5944	CGGCGGAGGCGCGGGGTGTCTCGGCCACGCGCGGTCTTAAGGCGATCGAGCGGACCGTGG	6003					
QY	469	GCTTTCCTGTGACTGGGATTCGACTTCATGACCGCGCTGACCTTGCGAAACCGGCTGCG	528					
Db	6004	GCGTTCAAGGAGATGGGTTTCGACTGCTGACCGCGGTGCAACTGCGCAACCGGCTGAAT	6063					
QY	529	GCGGTGACCGGGGTTCGGGTGGCGCAACATGTCTTTGACACACCGGACAGTGGACCGC	588					
Db	6064	GCGCGGACCGGGCTGCGGCTCCCGCGACAGTGTCTTTGACCAACCGAATTCGGGCTCTT	6123					
QY	589	CTCACCGCGC	598					
Db	6124	CTCACCGAGC	6133					

RESULT 3  
 US-11-056-470-2  
 Sequence 2, Application US/11056470  
 Publication No. US20050261263A1  
 GENERAL INFORMATION:  
 APPLICANT: SANTI, Daniel V.  
 APPLICANT: MYLES, David C.  
 APPLICANT: TIAN, Zong-Qiang  
 APPLICANT: HUTCHINSON, C. Richard  
 APPLICANT: JOHNSON, Robert G., Jr.  
 APPLICANT: ZHOU, Yi-Qiang  
 APPLICANT: FENG, Li  
 TITLE OF INVENTION: BENZOQUINONE ANSAMBYCINS  
 FILE REFERENCE: 300622007510  
 CURRENT APPLICATION NUMBER: US/11/056,470  
 PRIORITY FILING DATE: 2005-02-11  
 PRIOR APPLICATION NUMBER: US 10/212,962  
 PRIOR FILING DATE: 2002-08-05  
 PRIOR APPLICATION NUMBER: US 60/310,779  
 PRIOR FILING DATE: 2001-08-06  
 PRIOR APPLICATION NUMBER: US 60/389,255  
 PRIOR FILING DATE: 2002-06-14  
 PRIOR APPLICATION NUMBER: US 60/393,929  
 PRIOR FILING DATE: 2002-07-03  
 PRIOR APPLICATION NUMBER: US 60/395,275

;; PRIOR FILING DATE: 2002-07-12  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: FASTSEQ for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 6360  
;; TYPE: DNA  
;; ORGANISM: Streptomyces geldanus  
US-11-056-470-2

Query Match 22.0%; Score 132; DB 14; Length 6360;  
Best Local Similarity 55.2%; Pred. No. 4,6e-20;  
Matches 327; Conservative 0; Mismatches 250; Indels 15; Gaps 3;

QY 1 GCCTCTACGCGGCGGCAAGCGCTTCTCTGACGCTTCTGCGCGCGCTGCTGCGCGAGT 60  
DB 5656 GGGACATCTCTCCCGCCCAAGCGCTTCTGACGCTTCTGCGCGCGCTGCTGCGAG 5715  
QY 61 GGGCTGCGGCTCACTCTGATCGCTGCGGCTCTGCGCGCGCGAGAA---CATGGCCGCT 117  
DB 5716 GGGCTCCCGCCCACTGCTGCGGCTGCGGCTGCGCGCGAGAACAGTGGCATGAGCGCC 5775  
QY 118 ACCGAGGCGGCGCATCTGCTGCGCGCGCGCTGCGCGCATGAGACCGCGAGCGGCG 177  
DB 5776 GGGCTGCGGCTCGCGCGATGCGACCGGCTGCGCGCGCGCATGCTGCGGCTGCTGCG 5835  
QY 178 ATGAGAGGCTGCGGAGCAACCACTGAGACGCGGAGAACCGGCGGCTGCGGCTGCG 237  
DB 5836 GCGGAGGCGCTGAGACCTGCTTGAACCGGCTGCGGCTGCGGCTGCTGCTGCTGCG 5895  
QY 238 GACCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 296  
DB 5896 TTGCGGCTGCGACATCGCGCGCTGCGGCGGAGCGCGCGAGACCGCGCGCTGCTG 5955  
QY 297 ACTCGTGGGCTCGCGCGCGCGG-----CGAGAGACCGGCTGAGGATCGGAT 345  
DB 5956 GGTCTGCTCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6015  
QY 346 CTGCGCGCGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405  
DB 6016 CTGCGCGAGACGCTGCGCGCGCGCTGCTGCGCGCGAGACGAGCGCTGCTGCTG 6075  
QY 406 GTCCGAGCGCGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465  
DB 6076 ATCCGAGACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6135  
QY 466 GTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 525  
DB 6136 CAGGCGCTTCTGCGCGAGATGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 6195  
QY 526 GCGGCGGCTGACCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 577  
DB 6196 AACACGCGCGCGCGCTGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6247

## RESULT 4

US-11-075-185-37  
; Sequence 37, Application US/11075185  
; Publication No. US20050266434A1  
; GENERAL INFORMATION:  
; APPLICANT: REEVES, CHRISTOPHER D  
; APPLICANT: JULIEN, BRYAN  
; APPLICANT: REID, RALPH  
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBUTICINS  
; FILE REFERENCE: 010099.03  
; CURRENT APPLICATION NUMBER: US/11/075,185  
; PRIOR FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/551,103  
; PRIOR FILING DATE: 2004-03-08  
; PRIOR APPLICATION NUMBER: US 60/568,290  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 37

;; LENGTH: 10524  
;; TYPE: DNA  
;; ORGANISM: Sorangium cellulosum  
US-11-075-185-37

Query Match 16.8%; Score 101.2; DB 14; Length 10524;  
Best Local Similarity 59.3%; Pred. No. 1,4e-13;  
Matches 172; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 305 GGGTCCGCGCGCGCGCGCGAGAGACCGGCTGAGAAATCGATCTGCGCGCGCTGCG 364  
DB 5039 GCGCGCGCGCGCGCGCGAGCGCGCGAGCGCGAGCGCTCGCGCGCTGCGCGCGCT 5098  
QY 365 CGATGCGGAGCGCGCGAGACGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGAG 424  
DB 5099 GACGCGCGCGCGCGAGACGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5158  
QY 425 CGGTGCTGCG 484  
DB 5159 GCGTCTGCG 5218  
QY 485 GATTGCACTTCATGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 544  
DB 5219 GGTCTGCACTTCTGCACTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 5278  
QY 545 GGGTGGCGACGACGATCTGCTTGAACACCGCGCGCGCGCGCGCGCGCGCGCG 594  
DB 5279 AGCAGCGCGCGCGCGCGCGCTGCTTGAACATTCGCGCGCGCGCGCGCGCGCG 5328

## RESULT 5

US-11-075-185-1  
; Sequence 1, Application US/11075185  
; Publication No. US20050266434A1  
; GENERAL INFORMATION:  
; APPLICANT: REEVES, CHRISTOPHER D  
; APPLICANT: JULIEN, BRYAN  
; APPLICANT: REID, RALPH  
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBUTICINS  
; FILE REFERENCE: 010099.03  
; CURRENT APPLICATION NUMBER: US/11/075,185  
; PRIOR FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/551,103  
; PRIOR FILING DATE: 2004-03-08  
; PRIOR APPLICATION NUMBER: US 60/568,290  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1  
; LENGTH: 78869  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-11-075-185-1

Query Match 16.8%; Score 101.2; DB 14; Length 78869;

Best Local Similarity 59.3%; Pred. No. 1,4e-13;  
Matches 172; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 305 GGGTCCGCGCGCGCGCGCGAGAGACCGGCTGAGAAATCGATCTGCGCGCGCTGCG 364  
DB 36945 GCGCGCGCGCGCGCGCGAGCGCGCGAGCGCGAGCGCTCGCGCGCGCGCTGCG 37004  
QY 365 CGATGCGGAGCGCGCGAGACGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGAG 424  
DB 37005 GACGCGCGCGCGCGAGACGATGCTGCGCGCGCTGCTGAGCTGCTGAGGCGCG 37064  
QY 425 CGGTGCTGCG 484  
DB 37065 GCGTCTGCG 37124  
QY 485 GATTGCACTTCATGACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 544  
DB 37125 GGTCTGCACTTCTGCACTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 37184







Qy	445	CCGAGCGGTCATCGACGCTGAGCTGCTTCCGTGACCTGGGATTTGACATCCATGACCGC	504
Db	5200	CCGCGCAGAGTCCCCCGCGCAAGCCGCTGAAGAGCTGGGGCTCACTCGCTCAGCGCC	5259
Qy	505	GTGCACCTGGGAAACCGGCTGCGCGCGGTGACCGGGGTCGGGGTGGCAGACCATGCTC	564
Db	5260	GTGCAGCTGAGAAACCGCTTGGCGGCGCGCGCGCAGACGCGGTGGCGGGCAGCCTCGCG	5319
Qy	565	TTGCAGCACCCGAC	578
Db	5320	TTGCAGCATCCGAC	5353

```

RESULT 10
US-10-522-037-2
; Sequence 2, Application US/10522037
; Publication No. US20050282166A1
; GENERAL INFORMATION:
; APPLICANT: LIBRAGEN
; TITLE OF INVENTION: Method for the expression of unknown environmental DNA into adapt
; TITLE OF INVENTION: cells.
; FILE REFERENCE: B0149W0
; CURRENT APPLICATION NUMBER: US/10/522,037
; CURRENT FILING DATE: 2005-01-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO. 2
; LENGTH: 37507
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: DNA sequence of clone FS3-135.
US-10-522-037-2

```

Query Match	Similarity	11.9%	Score 71.8	DB 8	Length 37507
Best Local	Similarity	48.4%	Pred. No. 3.8e-07		
Matches	295	Conservative	0	Mismatches 302	Indels 12
				Gaps	3
QY	1	GCCTCTACGCGCGCGGCAACGCTTCTCTCGACGCTTTCGCCCGTCGTGTGCGACGAGT	60		
Db	13270	GCCAACTACATCGCGGCGCAACTGTCTCTCGAGGCACTCCGCAAGGCGCGCGCTCGAG	13322		
QY	61	GGGCTGCCGTGTACCTCGATGCTCTGGG---TCTGTGGGCGCGGCGAAGCATGGC----	113		
Db	13330	GGCCAGCGCGGCTGTGTGTGGGCTGTGGGCGCATCGGGGACGTGGCTACCTTCGCGCG	13389		
QY	114	--CGGTACGAGGGCGCGCACTACCTTGCGAGCGAGGCGCTGCGCGCATGGAACCGGAG	171		
Db	13330	CACGAGAAAGTGAAGAGAAAGCCCTTGCAACGCAATGGGCGGACCGGACTCGATCGGAG	13444		
QY	172	CGGGGATCGAGAGCTGCGGACCACTCTGAGCGCGCGGGAACCGTGTGGTGTGCGTGTG	231		
Db	13450	GCCGGCTCGCGGTGTGAGACAGACTCTGTGTCCCGACCTCTCGGCGCTTCGCGTGTG	13505		
QY	232	GACCTGGAACCGGAGCGGTTCTGTGAACTGTTCACCGCGCGCGCGCGCCCTTCTTC	291		
Db	13510	GACTTCAGCTGCGGCACTTCGCGCGCTTCTTCCCAACCGCGCGCTTCGCGCGCTTCGCG	13566		
QY	292	GACGAACCTCGGTGGGTTCGCGCGCGGAGCGGAGGACCGGTTCAGAACTTCCTGCG	351		
Db	13570	GAGCTGGGCGCGCGCGGAGAGAGACGCGGAGGACGACGACGCGCGCGCGGAGCTGTGCG	13622		
QY	352	CGGCGGCTGGGCGGTGATGTCGCGGAGGCGGAAAGTCAAGCATGTGTGCGCGGCTGTCCGA	411		
Db	13630	CGGCTTCGCGCAAGAGCTGAAGCGCTTCGAGCTTCGCGGATGTTCACTCGACTGTGCGCG	13688		
QY	412	GCGGAGTGTGAGACGCGTGTCTGTGGGCAACGGACAGCGCGGACCGGTATCGAGCTGTGAC	471		
Db	13650	CGGAGAGTGGGCGGAGATCTTG---CGATTCGCGCGGACCGGCTTCGATATAGCGCGGACCC	13744		
QY	472	TTCGTGACCTGGAGTTGACTCCATGACCGCGCGTCACTGTGCGAAACGGCTTCGCGCG	531		

Db	13747	TTTGAGAGAAATGGGCAATGATTCGCTGATAGGGCGTGAAGCTGTGACCCGCGTGAGGGCG	13806
Oy	532	GTGACCGGGGGTCCGGGGTGGCAAGCAACATGCTTTTGACACACCGGACATGGACCGCGTC	591
Db	13807	CGCTTCGGCGTGAACCTGCGCGGTGATGTGCGCTCTCCGACGACCGCTGATTCGAGAAAGCTG	13866
Oy	592	ACCGCGGCAC	600
Db	13867	GTGACCGCGC	13875

```

RESULT 11
US-11-075-185-33
; Sequence 33, Application US/11075185
; Publication NO. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, CHRISTOPHER
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHTIC GENE CLUSTER FOR AMBRUTICINIDS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO. 33
; LENGTH: 3381
;
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-33

```

Query Match	Similarity	11.8%	Score 70.8	DB 14	Length 3381
Best Local	Similarity	48.8%	Pred. No. 9.8e-07		
Matches	251	Conservative	0	Mismatches	257
				Indels	6
				Gaps	2
QY	68	CGGTCACTCGATGCGCTGGGCTGTGGGCGGGCGAACAATGCGCGGTACCGAGGCG	127		
Db	2687	CGCACCCGAGTCTGCGCGTCCGCGCTGACCGGCGGTGCGCGCGACGCGAGGCGCTGTGG	2746		
QY	128	GGCACTACCTGCGGAGCGCAGGCGCTGGCGCGCATGACCCGCGAGCGGCGCATCGAGGAGC	187		
Db	2747	TCGGAGGCTTCACCGCGACGACGCGCGGACTTGGCGCGGCTGTGGCGCGCTGGCGCGC	2806		
QY	188	TGCGGACCAACCTTGACCGCGCGGAGACCGGTGGGTGTGAGTGTGGAACCTTGACCGGAGC	247		
Db	2807	TGCATGTGNCAGAGGCGCAGCGGTGCAATTGGCGCGCGGTGTGCGCCGTTGGCGCGGAGC	2866		
QY	248	GGTTGTGTGAATGTTTACCGCGCGCGCGCGCGCGCGCTTCTTGAAGAACTGGTGGGG	307		
Db	2867	TGTGTAACTTGCACGACGTACGCAATTCACGCGCGCGCTACTGTGTTGATACCAAGAA	2926		
QY	308	TCCGCGCGGGAGCGGAGGAGACCGG---TCAGGAATCGAATCTGCGCGCGCTGGCGT	364		
Db	2927	GGTTTGGCGCTCGCAGCGCGCGTTCAGCAATTGGGGAAGTGGTGTACAGAAAGCTGGCG	2986		
QY	365	CGAATCCGGAAGCGCAACTGACAGACATGTGCGCCCGGCTGTCTCGAGCGGAGGTGGCAG	424		
Db	2987	GGCTGTCTCCCGCGCGGAGGAAACCGCTGCTCGAATGGGATCGCGCAAAATCCCGG	3046		
QY	425	CGGTGTCTGGGCGACGCGACCGCGGACGCGATCGAGCGTGAAGTGGCGCTTCGTAACCTGG	484		
Db	3047	CGGTGTCTGGGCGTGGAGCGCGCGGGGTG---CGCGCGGAGTTCGCTGGCGGATCTCG	3103		
QY	485	GATTCGACTTCATGACCGCGGTTCGACCTGCGGAAACCGGCTGCGCGCGGTGACCGGAGTCC	544		
Db	3104	GGTTAATTCGCGCATTCGCGGTGGAGCTGGGGAAGCGCGCTTGAGACGACGACGACGCGGA	3163		
QY	545	GGGTGGCCACGACCATGCTTTCACCAACCCGAC	578		

Db 3164 AGCTGCCGTCTCTGTTACACGACCCGAC 3197

RESULT 12

US-11-254-686-11

Sequence 11, Application US/11254686

Publication No. US20060040877A1

GENERAL INFORMATION:

APPLICANT: Martin, Christine J.

APPLICANT: Burns, Lesley S.

APPLICANT: Graupner, Paul R.

APPLICANT: Lewer, Paul

APPLICANT: Vousden, William A.

APPLICANT: Wilkinson, Bartie

APPLICANT: Waldron, Clive

TITLE OF INVENTION: Novel Spinosyn-Producing Polyketide Synthases

FILE REFERENCE: 62069

CURRENT APPLICATION NUMBER: US/11/254,686

CURRENT FILING DATE: 2005-10-20

PRIOR APPLICATION NUMBER: US 60/358075

PRIOR FILING DATE: 2002-02-19

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 11

LENGTH: 1694

TYPE: DNA

ORGANISM: Streptomyces avermitilis

US-11-254-686-11

Query Match 11.7%; Score 70.2; DB 11; Length 1694;

Best Local Similarity 49.9%; Pred. No. 1.5e-06;

Matches 177; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 246 GCGGTCGTGAACTGTTACCGCCCGCCCGCCCTCTTGACAACTCGGTGG 305

Db 1239 GCTGCCCGGCGGCGACGCGACCTTGACCTTCTGACGCTCTGCGCACTACGAGG 1298

Qy 306 GGTCCGCGCGGCGGCGGACGCGGTCAGAAATGGAATCGCCCGCGGCGGTGTC 355

Db 1299 GCAATGATGAGACTGCGGACCAACCGCTCTCGGACGCGCTTAACCGGCTGTCGCCAC 1358

Qy 366 GATCCGGAAGCGCGAAAGTCACGAGATGTCGCCGCGTGTGTCGAGCCGAGGTGCGAGC 425

Db 1359 GAGAGGACAGTGTCTGCTCTGACCTGTGTGCGCGGCCACACCAATGCGGTGCTGAACGA 1418

Qy 426 GGTGCTGGGCGACGCGACCGCGAGCGGTATGAGCGTGAAGCTGCGCTTCGATGCTGGG 485

Db 1419 CGACGGAAAGACGCGACCGCGGTGATGCGCGGCCCATGCGGAGTTTGCCCACTCGG 1478

Qy 486 ATTGCACTGCATGACCGCGCGTGTGACCTGCGGAAACCGGCTGCGCGCGGTGACCGGGGTCCG 545

Db 1479 CTTCGACTCCGTATGAGGTGTGAACTGCGGAAACCGCTTCAAGAAAGCGCCACGCGCTGCG 1538

Qy 546 GGTGCGCAGACCATGCTCTTCGACCAACCGGACGATGAGACCGGCTGACCGCGGAC 600

Db 1539 GTTGCGCGGAGCGCTCATCTTGAACACACGCGCGCGCGGTGCGCGCGCGC 1593

RESULT 13

US-11-175-689-2

Sequence 2, Application US/1175689

Publication No. US20060024806A1

GENERAL INFORMATION:

APPLICANT: STINEAR, TIMOTHY P.

APPLICANT: COLE, STEWART T.

APPLICANT: LEADLAY, PETER F.

APPLICANT: SMALL, PAMELA L.C.

APPLICANT: JOHNSON, PAUL D.R.

APPLICANT: JENKIN, GRANT A.

APPLICANT: DAVIES, JOHN K.

APPLICANT: HAYDOCK, STEPHEN F.

TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING

TITLE OF INVENTION: NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES

FILE REFERENCE: 03495.0329-01

CURRENT APPLICATION NUMBER: US/11/175,689

CURRENT FILING DATE: 2005-07-07

PRIOR APPLICATION NUMBER: 10/987,592

PRIOR FILING DATE: 2004-11-15

PRIOR APPLICATION NUMBER: 60/519,864

PRIOR FILING DATE: 2003-11-14

NUMBER OF SEQ ID NOS: 36

SOFTWARE: PatentIn Ver. 3.3

SEQ ID NO 2

LENGTH: 7233

TYPE: DNA

ORGANISM: Mycobacterium ulcerans

US-11-175-689-2

Query Match 11.6%; Score 69.6; DB 14; Length 7233;

Best Local Similarity 52.8%; Pred. No. 1.5e-06;

Matches 150; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 317 GGGCCGAGAGACCGGTCAAGATTCGATTCGCGCGCGGCTGCGAGCGGAGCGGTGCGAGG 376

Db 6110 GCGCGCGGCGGCTTGTCCGCAACGATTCGCTGCGCCGCTCAACGACTTAGCCCC 6169

Qy 377 CCGAAGTCACAGCATGTGCGCCGCGCTGTCGAGCGGAGCGGTGCGAGGCGGCGG 436

Db 6170 AAGAGCAACAAACACTGCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6229

Qy 437 ACGGACGCGCGAGCGGTATGAGCGTGAAGTCCCTTCGTCGACTGGAATTCGACTCA 496

Db 6230 ACCAGACCGCGCGAGGACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6289

Qy 497 TGAACCGCGTCACTGCGGAAACCGGCTGCGCGCGGCGGTGACCGGCGGTGCGGCGCA 556

Db 6290 TGAACCGCGTCACTGCGGAAACCGGCTGCGCGCGGCGGTGACCGGCGGTGCGGCGCA 6349

Qy 557 CCATGCTTCGACGACCGGACGATGAGACCGGCTGACCGGCGGAC 600

Db 6350 CCTCATCTTCATGATCACCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6393

RESULT 14

US-10-802-796-658/c

Sequence 658, Application US/10802796

Publication No. US20050250104A1

GENERAL INFORMATION:

APPLICANT: COLE, STEWART

APPLICANT: BUCHRIESEN-BROSCH, ROLAND

APPLICANT: GORDON, STEPHEN

APPLICANT: BILLAULT, ALAIN

TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST

TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED

TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF

TITLE OF INVENTION: MYCOBACTERIA.

FILE REFERENCE: 05394.0011-00000

CURRENT APPLICATION NUMBER: US/10/802,796

CURRENT FILING DATE: 2004-03-18

PRIOR APPLICATION NUMBER: US/09/673,476

PRIOR FILING DATE: 2002-03-29

PRIOR APPLICATION NUMBER: PCT/IB99/00740

PRIOR FILING DATE: 1999-04-16

PRIOR APPLICATION NUMBER: 09/060,756

PRIOR FILING DATE: 1998-04-16

NUMBER OF SEQ ID NOS: 743

SOFTWARE: PatentIn Ver. 2.2

SEQ ID NO 658

LENGTH: 444

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

NAME/KEY: modified\_base

LOCATION: (296)

OTHER INFORMATION: a, t, c or g

```

? NAME/KEY: modified_base
? LOCATION: (314)
? OTHER INFORMATION: a, t, c o r g
? FEATURE:
? NAME/KEY: modified_base
? LOCATION: (367)
? OTHER INFORMATION: a, t, c o r g
? FEATURE:
? NAME/KEY: modified_base
? LOCATION: (377)
? OTHER INFORMATION: a, t, c o r g
? FEATURE:
? NAME/KEY: modified_base
? LOCATION: (394)
? OTHER INFORMATION: a, t, c o r g
? FEATURE:
? NAME/KEY: modified_base
? LOCATION: (410)
? OTHER INFORMATION: a, t, c o r g
US-10-802-796-658

```

Every Match	11.4%	Score 68.4;	DB 8;	Length 444;
Best Local Similarity	53.1%;	Pred. No. 4.8e-06;		
Matches 138;	Conservative 0;	Mismatches 122;	Indels 0;	Gaps 0;

QY	DB	QY	DB
340	CGAGATCTCGCCCGCGCTGCGCTGATGCGGAGAGCCGAACGTTCACGAGCATGTCGCC	412	TGNGGGTTTGCCGCCCGCGTGGCCCGGTCCGTTCCGGCGAGACGAAGAAATTCGCCCGCAGCGCATC
399		353	
QY	400	CGGCTGTGTCCGAGCCGAGGTGGGAGCGGTGCTGGGGCCACGGCACGCCACGGTGAATCCGAG	459
DB	352	GAGGTGGTATGTGAGACATGCCCGACGGGTGCTGGGGCGGAGACCGGCGCTCCCGATCTMGAC	293
QY	460	CGTGAAGTTCGCGCTTCGCGTACCGTGGGATTTCGATCTCATGACCGCGCGGTGCACCTGCGGAAC	519
DB	232	GCTGGCCAGGCGGTGTTCGCCGATTCGGGATTTAAATTCCTTGAGTCCCGTGGAGATACCTAAC	233
QY	530	CGGCTTCGCGGCGGTGAACCGGAGGTCCGGGTGGGCCACGACCATCGTCTTTCCACCCCGACA	579
DB	232	CGCTTAAACAGCTGTCAACCGCAGTAACGCTGCGCGGACCGCGCATCTTCGATCACCCACCC	173
QY	580	GTGGACCGGCTCAACCGCGCA	599
DB	172	CCGACCGAACTAGCCAGTA	153

```

RESULT 15
US-11-221-284-658/C
/ Sequence 658, Application US/11221284
/ Publication No. US20060063182A1
/ GENERAL INFORMATION:
/ APPLICANT: Cole, Stewart
/ APPLICANT: Buchrieser-Brosch, Roland
/ APPLICANT: Gordon, Stephen
/ APPLICANT: Billault, Alain
/ TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
/ TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
/ TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
/ FILE REFERENCE: 3495-0169
/ CURRENT APPLICATION NUMBER: US/11/221,284
/ CURRENT FILING DATE: 2005-09-08
/ PRIOR APPLICATION NUMBER: US/09/060,756
/ PRIOR FILING DATE: 1998-04-16
/ NUMBER OF SEQ ID NOS: 743
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 658
/
/ LENGTH: 444
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURES:
/ NAME/KEY: unsure
/ LOCATION: (various positions within the sequence)
/ OTHER INFORMATION: applicants are uncertain of bases designated as "n"

```

US-11-221-284-658

Query Match	11.4%;	Score 68.4;	DB 11;	Length 444;
Best Local Similarity	53.1%;	Pred. No. 4.8e-06;		
Matches 138; Conservative	0;	Mismatches 122;	Indels 0;	Gaps 0;

QY	340	TCGAGATTCGCGCCGGCGGCTGGCGTCGATGCGCGAGAGCCGAAAGTTCACAGAGCATGTCGCC	359
Db	412	TCNCGGATTGCGCGCCCGGNTGGCCGGTGGGTGGCGGAGCAGAGCAANTGCGCGCAGCGATC	353
QY	400	CGGCTGTGCCAGCCGAGGTGGCAGCGGTGCTGGGGCCACGGCAGCCGACCGGTGATTCGAG	459
Db	352	GAGGTGGTATGTGAGCATGCGCGACCGGTGCTGGGGCGGAGCGGCGCTGCGGACTTNGAC	293
QY	460	CGTGAAGTCGCGCTTCGCTGAGCCTGGGAAATTCAGATCCATGACCGCGCTGCACCTGCGGAAAC	519
Db	292	GCTGGCCAGGCGCTTTGGCCGATTCGGGATTTAATCTTGAAGTGGCGGTGGAGATACGTAAAC	233
QY	520	CGGCTCGCGGCGGTGACCGGGGTCGGGATGGCGACGACCATGTGTTTCGACACCCGAGCA	579
Db	232	CGCTTAACAGGCTGTGACCGCAGTAAAGCTGCGCGACCGCGAGTCTTCGATGACCCGACCC	173
QY	580	GTGAGACCGGCTCACCGCGCA	599
Db	172	CGGACCGAATGAGCCGATTA	153

Search completed: April 7, 2006, 03:50:14  
Job time : 228.527 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:29:15 ; Search time 2314.42 Seconds  
(without alignments)  
12149.485 Million cell updates/second

```
Title: US-10-611-442-2_COPY_1200_1800
Perfect score: 601
Sequence: 1 gcctcctacgcgcyggcaaa.....ggaccgcctcacgcgcact 601
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```

1:  gb_est1: *
2:  gb_est2: *
3:  gb_est3: *
4:  gb_hlc: *
5:  gb_est4: *
6:  gb_est5: *
7:  gb_est6: *
8:  gb_est7: *
9:  gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	74.2	12.3	925	10	CNS0072P	AL053013 Drosophila
C	2	73.6	12.2	932	10	CNS00912Q	AL066742 Drosophila
C	3	72.4	12.0	1088	10	AG075424	AL0675424 Pan trogl
C	4	71.8	11.9	935	10	CNS006XX	AL066051 Drosophila
C	5	70.2	11.7	1033	10	CG156677	CG156677 P051-4-D
C	6	70	11.6	925	10	CNS0091P	AL053013 Drosophila
C	7	70	11.6	1270	10	AG382572	AG382572 Mus musc
C	8	69.6	11.6	776	10	CNS010RY	AL099352 Drosophila
C	9	69.6	11.6	1003	5	BO641288	BO641288 AGENCOURT
C	10	68.8	11.4	1007	5	BO650254	BO650254 AGENCOURT
C	11	67.8	11.4	1009	10	CNS010EW	AL098882 Drosophila
C	12	67.8	11.3	886	9	AZ196424	AZ196424 SP 1032 A
C	13	66.4	11.0	908	10	CW949347	CW949347 TCE33.4 G
C	14	66	11.0	1547	10	AG371264	AG371264 Mus musc
C	15	65.4	10.9	907	2	BG948266	BG948266 1024012H1
C	16	65.2	10.8	932	10	CNS0072Q	AL066742 Drosophila
C	17	65	10.8	1151	7	CK210749	CK210749 FGAS02257
C	18	64.8	10.8	996	10	AG072414	AG072414 Pan trogl
C	19	64.8	10.8	998	10	AG081195	AG081195 Pan trogl
C	20	64.6	10.7	1569	10	AG341503	AG341503 Mus musc
C	21	64.2	10.7	1038	10	AG060254	AG060254 Pan trogl
C	22	64.2	10.7	1280	10	CL470317	CL470317 SAIL 140

C	23	64	10.6	977	10	CL507759	SMIL_786
C	24	64	10.6	1552	10	AG430101	AG430101 Mus musculus
C	25	63.6	10.6	1244	2	B8846745	B8846745 1024014H0
C	26	63.2	10.5	834	10	AG332512	Mus musculus
C	27	63.2	10.5	1201	10	AG435403	AG435403 Mus musculus
C	28	63	10.5	894	10	AG075627	AG075627 Pan. trogl.
C	29	63	10.5	996	5	BUS31181	BUS31181 AGNCOCURT
C	30	63	10.5	1101	10	CNS0178Y	AL108460 Drosophila
C	31	62.8	10.5	1203	10	CNS015Y4	AL106054 Drosophila
C	32	62.8	10.4	1060	5	BUS38693	BUS38693 AGNCOCURT
C	33	62.8	10.4	1317	10	AG430654	AG430654 Mus musculus
C	34	62.6	10.4	916	10	CW938719	CW938719 TcB18.4_E
C	35	62.6	10.4	2420	10	CL509247	CL509247 SMIL_811
C	36	62.2	10.3	989	5	B8892601	B8892601 AGNCOCURT
C	37	62.2	10.3	1450	10	AG429718	AG429718 Mus musculus
C	38	62	10.3	1359	10	CL507830	CL507830 SAIL_788
C	39	62	10.3	1661	10	AG435089	AG435089 Mus musculus
C	40	61.8	10.3	539	6	CA741271	CA741271 waiA_c.pko
C	41	61.8	10.3	839	10	CNS0049B	AL1054280 Drosophila
C	42	61.6	10.2	707	6	CA160031	CA160031 SCJFPR3C0
C	43	61.6	10.2	1157	3	BMS42947	BMS42947 AGNCOCURT
C	44	61.4	10.2	856	10	AG883445	AG883445 Oryza sativa
C	45	61.4	10.2	1039	10	CW918239	CW918239 BDCA456TRR

## ALIGNMENTS

RESULT 1	CNS00091P/c	LOCUS	DEFINITION
	CNS00091P	925 bp	DNA linear
			GSS 03-JUN-1999
			Drosophila melanogaster genome survey sequence TERT3 end of BAC #

fly), genomic survey sequence.

VERSION AL053013.1 GI:4934461

**SOURCE** *Drosophila melanogaster* (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Ephydroidae; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)

**TITLE**

**COMMENT**

COMMENT

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP) .  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Oosagawa and  
Aron Mammone in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCT-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and BGT libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

```
source
1..925
/organism="Proscophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_1b="RPCT-98"
/notes="end : 1873"
```

**ORIGIN**













Oy ATGAGCCGGATTACGAGAGGCGGGGAGGACTTACTTCGCGACAGCCAGGGCTTGGCGGACCATGACACCG 168  
 Db GCGGCGCCCGGGGGGGGGCGCGGGGGGGGGSSSCGGGGGGGCGGGGGGGGGGGGGGGGGCGCGCGSS 331  
 Oy CAGCGCGGCAATCGAGAGAGCTGCGAGACCAACCTTGACGCGGGGAGAACCGCTGGGTGTCCGGT 228  
 Db CGSCTGGGCGGGGGGGCGGGGGGGGGGGGGCGSCGSGGGGGGGGGGGGGGGGGCGGGGGGGGGGG 271  
 Oy GTGGAACCTTGACCCGGAGACGGTTTCGTGCAAACTGTTCACCGCCGCCGCCGCCGCCGCCCTC 288  
 Db GCGCGGSGCGSACCGCGGGGGGGGGGGGGGGGGGGGGCGSGCGSGGGGGGGGGGGGGGGGGGGGG 211  
 Oy TTCCAGCAAACTCCGGTGGGGGTCCGCGCCCGGGGGCGAGAGACCGGTCAGAGAAATCCGAAATCTC 348  
 Db GTTCCGCGGCGCTTCGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGG 151  
 Oy GCCCGGCGCGCTTGCGCTCGATGCGCGGAGGCCGGAACGTTCACGAGCATGTGCGCCCGCTGGTC 408  
 Db GCGCGCGGCGGGGGGKKSAGCCCGSGGGGCGGCTTGCGGAGCGGSGGCGSGSGCGSGGCGGCG 91  
 Oy CGAGCGGAGGTGGCAGCGGTCTTGGGCGCACCGGACACCGGACGGTAAATGAACTTGACCTGACCTC 468  
 Db GSGCGGGGCGCGGGSSSCGGGGGGGGGGGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGCG 31  
 Oy GCCTTCCGTAACCTGGAGTTGCATCTC 494  
 Db GGGGTCGTCGCGCGCGGCGCTTGGAGCTC 5

	B0641288	1003 bp	mRNA	linear	EST 15-JUL-2007
RESULT 9					
B0641288/c					
LOCUS					
DEFINITION	AGNC/CORT_8287023 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:6292423				
	5', mRNA sequence.				

ACCESSION	BQ641288
VERSION	BQ641288.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

REFERENCE	1 (bases 1 to 1003)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

## FEATURES

### Source

```

Location/Qualifiers
1..1003
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6292423"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH1_MG_43"
/note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAC(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA Synthesis Kit

```

ORIGIN

(Stratagene) and Superscript II RT (Life Technologies)

Note: this is a NIH\_MGC Library. | "

Query Match	11.6%	Score 69.6;	DB 5;	Length 1003;
Best Local Similarity	44.3%	Pred. No. 0.00078;		
Matches 232; Conservative	0;	Mismatches 289;	Indels 3;	Gaps 1;

[illegible]

RESULT 10	1007 bp	EST 15-JUL-2002
BO650254		
LOCUS		
DEFINITION	AGNC00RT_8296624 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6270566	
	5', mRNA sequence.	
ACCESSION	BO650254	
VERSION	BO650254.1	
KEYWORDS	GI:21774426	
SOURCE	EST.	
ORGANISM	Homo sapiens (human)	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 1007)	
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D.	

Tissue Procurement: CGAP (Stanford)  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCm2445 row: o column: 15  
 High quality sequence set: 243.

FEATURES  
source  
Location/Qualifiers  
1..1007  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6270566"  
/issue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_11b="NIH MGC 100"  
/note="Organ: Liver; Vector: pOTB; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 11.4%; Score 68.8; DB 5; Length 1007;  
Best Local Similarity 43.7%; Pred. No. 0.0011;  
Matches 248; Conservative 0; Mismatches 314; Indels 6; Gaps 1;  
QY 27 CCTCGACGCTTCCGCGCGTGTGCTGCGCGACAGTGGCTGCGCTCACTTCGATCGCTG 86  
DB 426 CCGGGGCCCG 485  
QY 87 GGGTCTGTGGGCG 146  
DB 486 GCG 545  
QY 147 GGGCGTGGCG 206  
DB 546 GGGCGTGGCG 605  
QY 207 CCGGACCG 266  
DB 606 GGGCGTGGCG 665  
QY 267 CCG 326  
DB 666 GCG 725  
QY 327 GACCGGTGAGGATTCGATTCG 386  
DB 726 GGGCG 785  
QY 387 CGAGCATGTGCG 446  
DB 786 C-----GCCCG 839  
QY 447 GACGGTGTGAGCGTGTGCG 506  
DB 840 NCGGAGCG 899  
QY 507 CGACCTGTGCG 566  
DB 900 CCG 959  
QY 567 CGACCAACCG 594  
DB 960 GGGCG 987

RESULT 11  
CNS010EM/1009 bp DNA linear GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BACN03P19 of Drosophila library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL098882  
VERSION AL098882.1 GI:5610493

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
GSS.  
Drosophila melanogaster (fruit fly)  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1009)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage;  
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web: www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11.

FEATURES

source  
Location/Qualifiers  
1..1009  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACN03P19"  
/clone\_11b="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : SP6"

ORIGIN

Query Match 11.4%; Score 68.8; DB 10; Length 1009;  
Best Local Similarity 33.5%; Pred. No. 0.0011;  
Matches 123; Conservative 77; Mismatches 167; Indels 0; Gaps 0;  
QY 46 CGTGTGCG 105  
DB 965 CCGGCMRARGSAVTCG 906  
QY 106 AACATGCG 165  
DB 905 CCSSGGCG 846  
QY 166 CCGGCG 225  
DB 845 GCGACCG 786  
QY 226 GTGTGAGCCTGACCG 285  
DB 785 CGCNSG 726  
QY 286 CTCTTGACGAACTCGTGTGCG 345  
DB 725 CCCCMCSCCG 666  
QY 346 CTGCG 405  
DB 665 SGGGSG 606  
QY 406 GTCCGAG 412  
DB 605 TTYBGKG 599

RESULT 12  
AZ196424  
LOCUS SP\_1032.AL.A05.T7A Strongylocentrotus purpuratus, purple sea  
DEFINITION urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
genomic clone Plate=1032 Col=9 Row=A, genomic survey sequence.  
ACCESSION AZ196424  
VERSION AZ196424.1 GI:8390247  
KEYWORDS GSS.

SOURCE	Strongylocentrotus purpuratus
ORGANISM	Strongylocentrotus purpuratus Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 886)
REFERENCE	Cameron,R.A., Mahairas,G., Raat,D.P., Martinez,P., Biondi,T.R., Smartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
PUBMED	10920195
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division Of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-1047 Email: acameron@caltech.edu Plate: 1032 row: A column: 9 Seq primer: T7 Class: BAC ends High quality sequence stop: 886.
FEATURES	location/Qualifiers
source	1..886 /organism="Strongylocentrotus purpuratus" /mol_type="genomic DNA" /db_xref="taxon:7668" /clone_plate="1032 Col=9 Row=A" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin_sperm genomic BAC library" /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"
ORIGIN	
Query Match	11.3%; Score 67.8; DB 9; Length 886;
Best Local Similarity	46.7%; Pred. No. 0.0018;
Matches	233; Conservative 0; Mismatches 265; Indels 1; Gaps 1
Db	54 GCGCAGTGGGCTGCCGTGCACCTTCATCGCTTGAGGACCAGACAATGAGC 113
Db	359 GGCGCGCGGGGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 418
Gy	114 CGGTACCGAAGGCGCGGAC-TACTGTCCGACGACAGGCGCTGCGCGCAATGACCCGACGC 172
Db	419 CCGGGGGGGCGGGCGCGCGCGGNCGGCGCGGCGCGCGCGCGCGCGCGCGCGCN 478
Gy	173 GGGCATGCGAGGAGCTGCGGACCAACCTTGAGCGCGGGGACCCGTGAGGTGCGTGGTAG 232
Db	479 GGGGGCGCGCGCGCGCGGGGGGGGGGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
Gy	223 ACCTGAACCGGAGCGGTTGTCGAACTGTTCAACCGCGCGCGCGCGCGCGCGCCCTCTTCG 292
Db	539 GGCGGCGCGGCGCGGAGCNGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGCGCGGCGG 598
Gy	293 ACGAACTCGGTGGGTTCGCGCGCGCGCGCGCGGAGAACCGGTCAAGAAATCGAATTCTGCCC 352
Db	599 NCGGCGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 658
Gy	353 GCGCGCTGGCGTTCGATGCGCGGAGCGCAAAGTCAAGAGATATCGCCCGGCTGGTCCAG 412
Db	659 GGGGGCGCGCGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 718
Gy	413 CCGAGGTGGCAGCGGTGCTGGGCGCACGCGACGCGCGAAGGTGATCGAGCGTGAAGTCGCTT 472
Db	719 CCGGCGCGCGNCGCGGGGCGCCCGGGGGCGCCCGGGCGGCGCGCGCGCGCGCGCGCGCG 778
Gy	473 TCCTGACCTGGGAATTGCACTCAATGACGCGCGGTGACCTTGAGGAAACGCGCTCGCGCGG 532
Db	779 GGGGGCGCGCCCGCGGGGCGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGG 838

[illegible]

[illegible][illegible]

phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match	Score	DB 2	Length
10.9%	65.4	2	907

Best Local Similarity 46.6%; Pred. NO. 0.0053;  
Matches 275; Conservative 0; Mismatches 312; Indels 3; Gaps 2;

[illegible]

Search completed: April 7, 2006, 14:04:03  
Job time : 2317.42 secs







```
Db 1380 GAGAGCTGCGGACCACTTGAACGCGGGGACCCGTGGTGTGATGAACTTGAC 1439
Qy 241 CGGAGAGGTTTCGTGAACTGTTTCAACGCGGCGCGCGGCGCGCTTTCGAGAACTC 300
Db 1440 CGGAGAGGTTTCGTGAACTGTTTCAACGCGGCGCGCGCGCTTTCGAGAACTC 1499
Qy 301 GGTGGGTTCGCGCGCGCGGCGGAGAGAACGGTATCGAATCTTCGCGCGGCTG 360
Db 1500 GGTGGGTTCGCGCGCGCGGCGGAGAGAACGGTATCGAATCTTCGCGCGGCTG 1559
Qy 361 GCGTGAATGCGGAGGCGGAACTGACGAGCATGTCCGCGCGCTGTTCCGAGCGAGTG 420
Db 1560 GCGTGAATGCGGAGGCGGAACTGACGAGCATGTCCGCGCGCTGTTCCGAGCGAGTG 1619
Qy 421 GCAAGCGTGTGCGGCGGCGGAGCGGACGGTATCGAGGTTGACGTCCTTCGAGAC 480
Db 1620 GCAAGCGTGTGCGGCGGCGGAGCGGACGGTATCGAGGTTGACGTCCTTCGAGAC 1679
Qy 481 CTGGGATTTCGACTCATGACCGCGGTGACCTTGCGGAAACCGGCTTCGCGGCGAG 540
Db 1680 CTGGGATTTCGACTCATGACCGCGGTGACCTTGCGGAAACCGGCTTCGCGGCGAG 1739
Qy 541 GTCCGGGTGCGCAAGCAACCATGTCTTGACCACTCGACAGTGAACCGCTTCAACGCGAC 600
Db 1740 GTCCGGGTGCGCAAGCAACCATGTCTTGACCACTCGACAGTGAACCGCTTCAACGCGAC 1799
Qy 601 T 601
Db 1800 T 1800
```

## RESULT 2

```
US-10-672-396-5
; Sequence 5, Application US/10672396
; Publication No. US2004016567A1
; GENERAL INFORMATION:
; APPLICANT: SANTI, DANIEL V.
; APPLICANT: REID, RALPH C.
; APPLICANT: KODUMAL, SARAH J.
; APPLICANT: JAYARAJ, SEBASTIAN
; TITLE OF INVENTION: SYNTHETIC GENES
; FILE REFERENCE: 300622010900
; CURRENT FILING DATE: US/10/672,396
; PRIOR APPLICATION NUMBER: 2003-09-26
; PRIOR FILING DATE: 2002-09-26
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 9510
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-10-672-396-5
```

```
Query Match 46.8%; Score 281; DB 7; Length 9510;
Best Local Similarity 66.7%; Pred. No. 5e-60;
Matches 401; Conservative 0; Mismatches 200; Indels 0; Gaps 0;
```

```
Qy 1 GCTTCCTACGCGGCGGAGCAAGCCTTCTCTGACGCGCTTTCGCGGTGTGCGGCAAGT 60
Db 3778 GCGTCCTACGCGGCGGAGCAAGCCTTCTCTGATGTTTCGACGCGCGCGAGATCTGAA 3837
Qy 61 GGGCTCGCGGTACCTCGATCGCTGCGGTCTGTGGGCGCGGAGAACATGCGCGTACC 120
Db 3838 GGGCGACCGGTACGAGATGATCGATGGGGGTGTGGCGCGGTCAAGAAATATGCGCGTAT 3897
Qy 121 GAGGCGGCGCACTACCTGCGGAGCGAGCGGCGCTGCGCGCAATGAGACCGGCGGATC 180
Db 3898 GAGGCGGCGCACTACCTGCGGAGCGAGCGGCGCTGCGCGCAATGAGACCGGCGGATG 3957
```

```
Qy 181 GAGAGCTGCGGACCACTTGAACGCGGGGACCCGTGGTGTGATGAACTTGAC 240
Db 3958 GAAGAACTGATATCAAGCTGATATCAAGCTGCAAGACCTCGTCTCAAGTGTGATATGAC 4017
Qy 241 CGGAGAGGTTTCGTGAACTGTTTCAACGCGGCGCGCGGCGCGCTTTCGAGAACTC 300
Db 4018 CGTGGCGTTTTCGTGAGTGTGTTCAAGCTGCGCGGTACGCGCTTTCGAGAACTC 4077
Qy 301 GGTGGGTTCGCGCGCGCGGCGGAGAGAACGGTATCGAATCTTCGCGCGGCTG 360
Db 4078 GGTGGGTTCGCGCGCGCGGCGGAGAGAACGGTATCGAATCTTCGCGCGGCTG 4137
Qy 361 GCGTGAATGCGGAGGCGGAACTGACGAGCATGTCCGCGCGCTGTTCCGAGCGAGTG 420
Db 4138 GCGTGAATGCGGAGGCGGAACTGACGAGCATGTCCGCGCGCTGTTCCGAGCGAGTG 4197
Qy 421 GCAAGCGTGTGCGGCGGCGGAGCGGACGGTATCGAGGTTGACGTCCTTCGAGAC 480
Db 4198 GCAAGCGTGTGCGGCGGCGGAGCGGACGGTATCGAGGTTGACGTCCTTCGAGAC 4257
Qy 481 CTGGGATTTCGACTCATGACCGCGGTGACCTTGCGGAAACCGGCTTCGCGGCGAG 540
Db 4258 CTGGGATTTCGACTCATGACCGCGGTGACCTTGCGGAAACCGGCTTCGCGGCGAG 4317
Qy 541 GTCCGGGTGCGCAAGCAACCATGTCTTGACCACTCGACAGTGAACCGCTTCAACGCGAC 600
Db 4318 GTACGTGAGGCTGCGCAAGTGTATTTGACCATCAAGATCAAGCGCTTTCGCGAGTAT 4377
Qy 601 T 601
Db 4378 T 4378
```

## RESULT 3

```
US-10-205-032-17
; Sequence 17, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Stefa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSABAMICIN
; FILE REFERENCE: 3016-205
; CURRENT FILING DATE: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 4725
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-17
```

```
Query Match 36.6%; Score 220.2; DB 5; Length 4725;
Best Local Similarity 61.6%; Pred. No. 5e-45;
Matches 370; Conservative 0; Mismatches 228; Indels 3; Gaps 1;
```

```
Qy 3 CTCCTACGCGGCGGAGCAAGCCTTCTCTGACGCGCTTTCGCGGTGTGCGGCAAGT 62
Db 3879 GCGTACGCGGCGGAGCAAGCCTTCTCTGACGCGCTTTCGCGGTGTGCGGCAAGT 3938
Qy 63 GCTGCGGTACCTCGATCGCTGCGGTCTGTGGGCGCGGAGAACATGCGCGTACC 122
Db 3939 GCTGACGCGCACTCGGTGCGGTCTGTGGGCGCTGCGGAGCAATGCGGAGGCGGA 3998
Qy 123 GGGCGGACATCTGCGGAGCGGAGCGGCTGCGGCAATGAGACCGGCGGCGATGGA 182
Db 3999 GCGCGAGAGTATCTGAGCGGCGGCGGCTGCGGCGGATGCTCCGAGCGTGGCGTGA 4058
Qy 183 GGAAGTGGAGCAACCTTGAACGCGCGGAGAACCGGTGGGTGTGCGGTGTGAGACTGACG 242
Db 4059 CGCCCTCTGCGCGCGCTGAGCGGAGCAAGCTTTCGTGCGCGTTCGCGGAGTGAATG 4118
```

OY 243 GGAGGGGTTGCGTGAACGTGTCACCGCGCGCCCGCGCGCGGCCCTTTCGACGAACTCGG 302  
 DB 4119 GACGCTGTTCACGCGCGGGTTTACCGCGCTTCCGCGCCGACCGCGCTGCTCGGGACTTCC 4178  
 OY 303 TGGGGTCCGCGCGGGGGCGGAGAGACCGGTC--AGGAATTCGATCTTCGCGCGCGGCT 359  
 DB 4179 GGAGGCGCGCGGAGCGTGGCGGAGCGCGGAGCCGCGGGGCTTCGACTGCGCGGCTTGGCA 4238  
 OY 360 GGCGTTCATGCCGAGCGCCGAAAGCTTCACGAGCATGTCCGCGGCTGTTCGAGCCGAGGT 419  
 DB 4239 CGCGCGCGCGAGCCCGAAGAACGCGCGCGGGGCTGTCTCGACTGTGATCCCGCGCAGGT 4238  
 OY 420 GGCAAGCGGTGCTGGGCGACGCGACGCGAGCGGTATTCGAGCGTGAAGTGTGCGCTTCCGGTA 479  
 DB 4239 CGCGCGCGCTCTCGGCGACCGGGGGCCGAGGACGTCGCGGCCCGAAGCGCGGTTCCGGGA 4358  
 OY 480 CCTGGGATTCGACTTCATGACCGCGCGTGAAGCTGTGCGGAAACGCGCTCGCGCGGTTGACCGG 539  
 DB 4359 GATCGGATTCGACTGTGAGCGCGCGTGAAGCTGTGCGGAAAGGCGCTCAAGGGCGCGCGTGG 4418  
 OY 540 GGTCCGGGTGGCGACGACCATGCTGTTTGACCAACCGGACAGTGGACCGGCTTCAACCGGCA 599  
 DB 4419 CGTGCAGCTGTCCGCAACCTCTGTCTTGACACCAACCCGACCGCGAGCGGCTGCGCGAGCA 4478  
 OY 600 C 600  
 DB 4479 C 4479

```

RESULT 4
US-10-205-032-1
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Scalfa, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSABARBITIN
; FILE REFERENCE: 301.6-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1

```

Query Match	36.6%	Score	220.25	DB	5	Length	60196
Best Local Similarity	61.6%	Pred.	No. 3e-45				
Matches	370	Conservative	0	Mismatches	228	Indels	3
				Gaps	1		
Qy	3	CTCTTACGCGGGGCAACGCTTCTCTGACAGCCCTTGCCTGTGTGTGCGCGCAGATGG	62				
Db	40170	CGCCTTACGGGGCGGCGCAACGCGGCGCTCTGACAGCGCTGCCGAAACGGCGCGCGCCGCCGG	40222				
Qy	63	GCTTCCGGATCCACTTCGATTCGGCTTGGGGTCTGTGGGCGCGGCGAGAACATGGCCGATACCGA	122				
Db	40230	GCTTACCGCCACCTCTCGTCCGCTTGGGGCTGTGGGGCTCCGGGGGCAATGGGCCAGAGGCCAG	40285				
Qy	123	GGGCGCGACCTACTCTGCGCAGCCAGGGCTTGCAGCATGAACTCCGACGGCGGCGATTCGA	182				
Db	40290	CGCGAGAGATTAACCTTAGCCGCGCGGGGCTGCGCGCGCATCTCTCCGAGGTGCGCTGGA	40345				
Qy	183	GGAAGCTTGGACCAACCTTGAACGCGCGGGGACCCGATGGATGTTCGATGGATGAACCTTGACCG	242				
Db	40350	CGCCCTCTTGGCGGCGCTTGACCGGGACAGAGACTCTGTGCGCGTGCACACTTGACATG	40405				
Qy	243	GGAAGCGTGTTCGAACTGTTCACTGGCGCGCGCGCGCGCCCTCTTTCGACGAATCTCG	302				
Db	40410	GAGCTCTTTCACGCGCGGGATTCACCGGATTCGAGGCCAGGCGGCTGTCTGGCGCACTTCC	40465				

QY	303	TTGGGATCCGAGCCCGGGGCGGAGAACCGGCT---	AGAAATCGGATCTTGCCCGGCGGCT	359
Db	40470	GGAGGCCCGCGCAAGCTGGCCCAACCCCGA	CCCGGGGCTCCGACTGCCCGCTGGCA	40529
QY	360	GGCGTCGATCCGGAAGCGCAAGTCAACGAGCATG	TGCGCCGGCTGGTCCGACCGAGGT	419
Db	40530	CGCCGCGCGAAGCCCGGACGAAAGCCGCGCGGAGCT	GTCTGACTGGTGTACCGCGCAGGT	40589
QY	420	GGCAGCGGTCTGGGCCACGCGCACGCGTGA	TGACGCTGACGTCGCTTCGTGA	479
Db	40590	CGCGCGGCTCTCGGCACACCGGGGGCCGAGCA	CGTCGGCCCGACGCGCGCTTCGGGA	40649
QY	480	CTTGGGATTGACTCCATGAACGCGCGTGA	CTCTGCGGAAACCGGCTTCGCGCGGTGA	539
Db	40650	GATCGAATTCGACTCGGTGACCGCCCTCTCA	ACTGTGGCAAGCGGCTCAGGGCGCGGTGCG	40709
QY	540	GGTCGGGATGGCAACGACATCGTCTTTGAC	CAACCCGACAGTGAACCGGCTCAACGGCGCA	599
Db	40710	CGTACCGCTGTCCGCCACCTCTGTTCGAC	CAACCCACCGCAGCGCGGTCGCGGAGCA	40769
QY	600	C	600	
Db	40770	C	40770	

```

RESULT 5
US-10-156-761-2886
: Sequence 2886, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMODA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: PRIOR FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 2886
:
: LENGTH: 18438
: TYPE: DNA
: ORGANISM: Streptomyces avermitilis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(18438)
US-10-156-761-2886

```

	Query Match	34.9%	Score 210	DB 6	Length 18438
	Best Local Similarity	61.3%	Prod. No. 1	2e-42	
	Matches	350	Conservative	0	Mismatches 220; Indels 6; Gaps 1;
Qy	1	GCTCTCTACGGGGGCGCAACGCTTCTCTGACGCTTTCGCCCTGTGTGGTTCGGCGCACT	60		
Db	11284	GGGGCGTATACGGCCCGCCCAACGGTATCGCCGACGCCCTCGCGCGACCGCGCCACAGGCC	11341		
Qy	61	GGGCTGCGCGGTCACTCTGATTCGGCTTGAGGTCTGTGTGGCCCGCGCAGAACTAGCCCGGTACC	120		
Db	11344	GGCTTCACCGCCACTCTCCGTGGGCTTGAGGCACTGTGGCCCGCGCGGAGATGGTTCAGCAGC	11403		
Qy	121	GAGGGCGGGCGACATACCTGTGGCGACCGGAGCGGCTGTGGCGGCATATGACCCCGAGAGGGCGGANC	180		
Db	11404	GACCTGGCGCGGGACCTTCGCCCGCGCGGGGTGTGCTTCATATGACCCCGACCGGGCGAATC	11465		
Qy	181	GAGGAGTGTGCGAACCAACCTTGAGCGCCGGGAGCCCGTGGGTGTGGTGTGTGACCTTGGAC	240		

Db 11464 GCGCGCTCCAGCAGGCGCTTCGACGACGAGACCAAGCTGACGTCACCGATGAGC 11523  
Qy 241 CCGAGAGCGTTTCGTGCAACTGTTTCAACGCGCGCCCGCGCGCCCTCTTTCAGACGAACTC 300  
Db 11524 TGGAGAGCGTTTCGCGGAGAGTTTCAACGCGCGCCCGCGCGCCCTCTTTCAGACGAACTC 11583  
Qy 301 GGTGGAGTCCG 354  
Db 11584 CCGAGAGCG 11643  
Qy 355 CCGAGAGCGTTTCGTGCAACTGTTTCAACGCGCGCCCGCGCGCGCCCTCTTTCAGACGAACTC 414  
Db 11644 AGGCTGCG 11703  
Qy 415 GAGGTGCGAGCGGTGCTGCG 474  
Db 11704 GCGCGCGCGCTTCG 11763  
Qy 475 CGTGAAGCGTTTCGTGCAACTGTTTCAACGCGCGCCCGCGCGCGCCCTCTTTCAGACGAACTC 534  
Db 11764 AAGGACCTGCG 11823  
Qy 535 ACCGAGGTCCGAGGTGCG 578  
Db 11824 ACCGAGGTACGTGCG 11867

## RESULT 6

US-10-156-761-15102  
Sequence 15102, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 15102  
LENGTH: 125746  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-15102

Query Match 34.9%; Score 210; DB 6; Length 125746;

Best Local Similarity 61.3%; Pred. No. 8.4e-43;

Matches 358; Conservative 0; Mismatches 220; Indels 6; Gaps 1;

Qy 1 GCTCTTACGCGCGGAGCAAGCTTCTTCGACGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 36594 GCGCGGTACG 36653  
Qy 61 GGGCTGCGCGTCACTCGATCG 120  
Db 36654 GGGCTGCGCGTCACTCGATCG 36713  
Qy 121 GAGGCGCGGCACTACCTGCG 180  
Db 36714 GAGCTGCGCGGCACTACCTGCG 36773  
Qy 181 GAGAGCTGCGGCACTACCTGCG 240  
Db 36774 GAGCGCTGCGGCACTACCTGCG 36833

Qy 241 CCGAGAGCGTTTCGTGCAACTGTTTCAACGCGCGCCCGCGCGCCCTCTTTCAGACGAACTC 300  
Db 36834 TGGAGAGCGTTTCGCGGAGAGTTTCAACGCGCGCCCGCGCGCCCTCTTTCAGACGAACTC 36893  
Qy 301 GGTGGAGTCCG 354  
Db 36894 CCGAGAGCG 36953  
Qy 355 CCGAGAGCGTTTCGTGCAACTGTTTCAACGCGCGCCCGCGCGCGCCCTCTTTCAGACGAACTC 414  
Db 36954 AGGCTGCG 37013  
Qy 415 GAGGTGCGAGCGGTGCTGCG 474  
Db 37014 GCGCGCGCGCTTCG 37073  
Qy 475 CGTGAAGCGTTTCGTGCAACTGTTTCAACGCGCGCCCGCGCGCGCCCTCTTTCAGACGAACTC 534  
Db 37074 AAGGACCTGCG 37133  
Qy 535 ACCGAGGTCCGAGGTGCG 578  
Db 37134 ACCGAGGTACGTGCG 37177

## RESULT 7

US-10-156-761-1/c  
Sequence 1, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 34.9%; Score 210; DB 6; Length 9025608;

Best Local Similarity 61.3%; Pred. No. 3.6e-43;

Matches 358; Conservative 0; Mismatches 220; Indels 6; Gaps 1;

Qy 1 GCTCTTACGCGCGGAGCAAGCTTCTTCGACGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 3615065 GCGCGGTACG 3615066  
Qy 61 GGGCTGCGCGTCACTCGATCG 120  
Db 3615065 GGGCTGCGCGTCACTCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3614946  
Qy 121 GAGGCGCGGCACTACCTGCG 180  
Db 3614945 GAGCTGCGCGGCACTACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3614886

QY 181 GAGAGCTGCGAACAACCTTGAACCGCGGAGACCCGTGGTGTCTGTGTGAACCTGAC 240  
Db 3614885 GCGCGCTCCAGCAGGCGCTTCCACACAGAGACCAAGCTGACGCTTCAACGACATGAC 3614826  
QY 241 GCGGAGCGGTTCCTGCAACTGTTTCAACCGCGCGCGCGCGCGCGCTTCTTCAACGAACTC 300  
Db 3614825 TGGAGGCTTCCTGCGGAGAGCTTCAACCGCGCGCGCGCGCGCGCTTCTTCAACGCAATC 3614766  
QY 301 GGTGGGCTCCG 354  
Db 3614765 CCGGAGCG 3614706  
QY 355 CGGCTGGCTGATGATGCGGAGCGCGCAAGTCAAGAGATGTCCGCGCGCGCGCGCGCG 414  
Db 3614705 AGGCTGGCGCGGATGATGCGGAGCGCGCAAGTCAAGAGATGTCCGCGCGCGCGCG 3614646  
QY 415 GAGGTGCGACCGGTGCTGCGCGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
Db 3614645 GCG 3614586  
QY 475 CGGAGCTGGGATTCGATGCTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 534  
Db 3614585 AAGGAGCTGGGCTTCGATGCTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3614526  
QY 535 ACCGCGGCTCCGCGGTGCGCAACGATGCTTTCGACGACCGCGCGCGCGCGCGCGCG 578  
Db 3614525 ACCGCGGCTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3614482

## RESULT 8

US-10-760-493-34  
/ Sequence 34, Application US/10760493  
/ Publication No. US20050187167A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Ecopia Biosciences Inc  
/ APPLICANT: Farnet, Chris M.  
/ APPLICANT: McAlpine, James B.  
/ APPLICANT: Zazopoulos, Emmanuel  
/ APPLICANT: Bachmann, Brian O.  
/ APPLICANT: Pirae, Mahmood  
/ TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE  
/ FILE REFERENCE: 3004-9US  
/ CURRENT APPLICATION NUMBER: US/10/760,493  
/ PRIOR FILING DATE: 2004-01-21  
/ PRIOR APPLICATION NUMBER: USN 60/441,123  
/ PRIOR FILING DATE: 2003-01-21  
/ PRIOR APPLICATION NUMBER: USN 60/469,810  
/ PRIOR FILING DATE: 2003-05-13  
/ PRIOR APPLICATION NUMBER: USN 60/491,516  
/ PRIOR FILING DATE: 2003-08-01  
/ PRIOR APPLICATION NUMBER: USN 60/494,568  
/ PRIOR FILING DATE: 2003-08-13  
/ NUMBER OF SEQ ID NOS: 78  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 34  
/ LENGTH: 9684  
/ TYPE: DNA  
/ ORGANISM: Streptomyces albusensis  
US-10-760-493-34

Query Match 34.5%; Score 207.6; DB 9; Length 9684;  
Best Local Similarity 60.3%; Pred. No. 5,4e-42;  
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

QY 1 GCGCTCTACGCGCGCGGCAACGCGCTTCTCTGACGCGCTTCCGCGCGCGCGCGCGCG 60  
Db 4054 GCGAATTAAGCGCGCGGCAACGCGCTTCTCTGACGCGCTTCCGCGCGCGCGCGCGCG 4113  
QY 61 GCGCTCTACGCGCGCGGCAACGCGCTTCTCTGACGCGCTTCCGCGCGCGCGCGCGCG 120  
Db 4114 GCGCTCTACGCGCGCGGCAACGCGCTTCTCTGACGCGCTTCCGCGCGCGCGCGCGCG 4173

QY 121 GAGGCGCGGCACTACCTGCGCAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 4174 GCGGTGTGTATCGGATGCG 4233  
QY 181 GAGAGCTGCGAACAACCTTGAACCGCGGAGACCCGTGGTGTCTGTGTGAACCTGAC 240  
Db 4234 GCGGCACTCAGACCGCGCTGACCGGAGACGACGACCGCTGACCGCTGACGACGCTG 4293  
QY 241 GCGGAGCGGTTCCTGCAACTGTTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 4294 TGGAGGCTTCCTGCGGAGAGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4353  
QY 301 GGTGGGCTCCG 336  
Db 4354 CCGGAGCGCGCGGAGGCTATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4413  
QY 337 GAATGAGATTCG 396  
Db 4414 GCGCGCGCGCTGCG 4473  
QY 397 GCGCGCGCTGCTCGGACCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 456  
Db 4474 CTGAACTGTGCTCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4533  
QY 457 GAGCGTAGTGTGCTTCGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 516  
Db 4534 GAGCGCGCGCGCGCGCTTCAAGAGCTGCGGCTTCACTTCACTTCACTTCACTTCACT 4593  
QY 517 AACCGGCTGCGCGCGGTGACCGGAGTCCGCGGTGCGCGCGCGCGCGCGCGCGCGCG 576  
Db 4594 AACCGGCTGAAAGCG 4653  
QY 577 ACAATGACCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
Db 4654 ACGCCACCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4677

## RESULT 9

US-10-760-493-18  
/ Sequence 18, Application US/10760493  
/ Publication No. US20050187167A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Ecopia Biosciences Inc  
/ APPLICANT: Farnet, Chris M.  
/ APPLICANT: McAlpine, James B.  
/ APPLICANT: Zazopoulos, Emmanuel  
/ APPLICANT: Bachmann, Brian O.  
/ APPLICANT: Pirae, Mahmood  
/ TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE  
/ FILE REFERENCE: 3004-9US  
/ CURRENT APPLICATION NUMBER: US/10/760,493  
/ PRIOR FILING DATE: 2004-01-21  
/ PRIOR APPLICATION NUMBER: USN 60/441,123  
/ PRIOR FILING DATE: 2003-01-21  
/ PRIOR APPLICATION NUMBER: USN 60/469,810  
/ PRIOR FILING DATE: 2003-05-13  
/ PRIOR APPLICATION NUMBER: USN 60/491,516  
/ PRIOR FILING DATE: 2003-08-01  
/ PRIOR APPLICATION NUMBER: USN 60/494,568  
/ PRIOR FILING DATE: 2003-08-13  
/ NUMBER OF SEQ ID NOS: 78  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 18  
/ LENGTH: 164051  
/ TYPE: DNA  
/ ORGANISM: Streptomyces albusensis  
US-10-760-493-18

Query Match 34.5%; Score 207.6; DB 9; Length 164051;  
Best Local Similarity 60.3%; Pred. No. 3.1e-42;  
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

```

QY      1  GCCTCTCAACGCGGCGGCAACGCGCTTCTCTGAGCGCTTGGCCCGTGTGTGGGCGAGT  60
Db      98110  GGCMAACTACCGCGCGCGGCAACGCGTTCCTGTGAAGCGCTTGGCGAGTACCGCGCGAGATCC  98169
QY      61  GGGGTGCGCGGTCACTTGATGAGCTGTGGGGGCTGTGTGGGCGCGGCAAGAACATGCGCGGTAC  120
Db      98170  GGACTGCGCGCGCACTTCTCATGTGCTGTGGGCGCTTGTGGGCGAGACGGGGGCAATGCGCGAGGCG  98229
QY      121  GAGGCGCGCGCACTTACCTGCGCAGCGCGGCGCTGCGCGCCATGAGACCCGACGCGGCGATC  180
Db      98230  GCGGTGCGGTACCGGAATGGCGCGCGCAAGGGGATCATGAGATGTGCGCGAGCGTCCGTC  98289
QY      181  GAGGAGCTGCGGACCACTTGAAGCGCGCGGAGACCCGTGGGTGTGTGTGTGATCCTTGAC  240
Db      98290  GCGCACTCTCAGCAGCGCGCTTGAAGCGGAGAGACGACCCCTGACCGTCGCGCGACATGAG  98349
QY      241  CGGAGAGCGGTTCGTGGAACCTGTTCACCGCGCGCGCGCGCGCGCGCGCTTGTGACGACATC  300
Db      98350  TGGAGAGCGCTTCTGTCTTGTGCGCTTCACTTCGCGCGCGCGCGCGCGCTGTGTCGACGACCTG  98409
QY      301  GGTGGGGGCGCG-----CGCGGGGCGAGAGACCGGTGAG  336
Db      98410  CCCGAGCGCGGAGGTCATGAGACGCCAGCGCAGAGCGCGGAGAGACCGCGAGC  98469
QY      337  GAATCGATCTCGCCCGCGCGCTGCGCTGCATGCGCGAGCGCGAACGTGACGAGCATGTC  396
Db      98470  GCGCGCGCGCTGTGCGCGCGAGCATGACCGCGCGCGCGCGAGCGGACGAGAGCGATGCTC  98529
QY      397  GCCCGGCTGTGTCGAGCGCGAGGTGCGCGGTGCTGGGCGACGCGCGACCGTGAATC  456
Db      98530  CTCGAACCTGTGTCCCAACCGCGGTGCGCGGTCTGTGCGCTTCAACCGGAGCGCGAGCGGTG  98589
QY      457  GAGGGTACGTCGCGCTTCCGTCGATGGAATTCGATTCATGACCGCGCGTCGACCTTGCGG  516
Db      98590  GAGGCGGCGCGCGGCTTTCAGAGAGAGTGGGCTTGAATCTCTTCACTTCCGTGAACTGCGC  98649
QY      517  AACCAGCTCGCGGCGGTGACCGGAGTCCGAGTGGCCAGCAGCATGTCCTTTCAGACCCG  576
Db      98650  AACCGCTGAAGCGGCGCAGCGGCTTAAGCTGCGCGCACCTCGTCTTTCAGACACCGG  98709
QY      577  ACAGTGACCGGCTTACCGCGGAC  600
Db      98710  ACGCCCAACGTCCTGCGCGCGGAC  98733

RESULT 10
US-10-205-032-19
; Sequence 19, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Stefka, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-2US
; CURRENT APPLICATION NUMBER: US/10/205,032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 5355
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-19

```

Query Match	34.0%	Score 204.4	DB 5	Length 5355
Best Local Similarity	60.0%	Pred. No. 3.7e-41		
Matches 359	Conservative	0	Mismatches 236	Indels 3
Gaps				1
Qy	3	CTCTAAGCGGCGGCAACGCTTCTCTCGACGCTTCGCGCTCGTGTGCGGCGAGTGG	62	
Db	3876	CGCGTACGCTGCGGCGGCAACGCGCGCTTCGACCGCGTGTGCGGCGCGCGAG	3935	

QY 63 GCTGCGGATCACTTCGATCGCTGGGGTCTGTGGGCGGGGAGAAATGGCCGGTACCGA 122  
 Db 3936 GCTGCGCGGACACTCGCTCGCTGGGGGGTCTGTGGGCGGGGCGCAATGGGGGCGGGCGC 3995  
 QY 123 GGGGCGGACCTACTCTGGGCGACCGGCGCTGGCGGCAATGGAACCGGAGGGGCGATCGA 182  
 Db 3996 GGGGAGAGATGACTCTGGCGCGCGCGCGGGCGCTCGGGGCATGCCCCCGGACGCGCTTCGC 4055  
 QY 183 GGAGCTGGCGGACCACTCTGAGCGCGGGGGAGACCGCTGGGGTGTGGTGGTGGACCTGGACCG 242  
 Db 4056 CGCCTTCGGGCGGATCTCTTGAGCGCGGACGAGAACCGGGGTGACGGTCTTCGACACCGAGTG 4115  
 QY 243 GGAGCGGTTCTGTGCAACTGTTCACCGCGCGCGCGCGCGCTTCCTTCGACGAACCTCGG 302  
 Db 4116 GGGCGGGTTCGCGTCCGGCTTCGCGCGCTTCGCGCGCGCGCTTCGCGCGCTGGCTGGCGCTGCC 4175  
 QY 303 TGGGGTCCGCGCGCGCGCGCGCGGACCGGTCAAGAAATCGGATTCGCGCGCGCGCTGGC 362  
 Db 4176 GGGCGGGGACCTCGATCGCGCGCGCGCGCGCGACCGGGCGGACGCGCGCGTGGCCCCGACC 4235  
 QY 353 GTTCGATGCGCGGAGGCGCGGACGTCAGAGGCAATGTGCGCGCGCTGGTTCGAGCGGAGTGGGC 422  
 Db 4236 GAGCCCGGACACCACTGACGCGCGCGCGCGCTGTGAGC---GTTGGTTCGCGCGCGAGGCGCGC 4292  
 QY 423 AGCGGTCTGGGGCACGAGCAGCGCGAGCGGTATGAGCGGTGACGTCGCTTCCTGTGACT 482  
 Db 4293 CGGGGTCTGGGGGACAGACGGTGTGCGACGACTTCGGCGGACCGGAGTTCTCGGCGCT 4352  
 QY 483 GGGATTGCACTTCATGACCGCGCTGTGACCTTGCGGAAACCGGCTCGCGCGGTGACCGGGGT 542  
 Db 4353 CGGCTTCGACTCGCTCGCGCGCGTTCAGACTGCGCGCGCGCTTCGCGGAGGACCACTCGGCT 4412  
 QY 543 CCGGGTGGCCAGACCACTGTCTTCGACCAACCGACGAGTGGACCGGCTCAACCGGAC 600  
 Db 4413 GAGCTCTTCGCGCGCGGTTCTGTTCAGACACCGACCCCTTACCGCGCTCGCGCGAC 4470

```

RESULT 11
US-10-937-730A-3
; Sequence 3, Application US/10937730A
; Publication No. US20050112726A1
; GENERAL INFORMATION:
; APPLICANT: HU, Zhihao
; APPLICANT: REID, Ralph
; TITLE OF INVENTION: Biosynthetic Gene Cluster for Leptomycin
; FILE REFERENCE: 010092.02
; CURRENT APPLICATION NUMBER: US/10/937,730A
; CURRENT FILING DATE: 2004-09-08
; PRIOR APPLICATION NUMBER: 60/502,423
; PRIOR FILING DATE: 2003-09-11
; PRIOR APPLICATION NUMBER: 60/553,384
; PRIOR FILING DATE: 2004-03-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 67167
; TYPE: DNA
; ORGANISM: Streptomyces sp. ATCC 39366
US-10-937-730A-3

```

Query Match	33.5%	Score 201.6	DB 9	Length 67187
Best Local Similarity	59.6%	Pred. No. 1.1e-40		
Matches 361	Conservative 0	Mismatches 239	Indels 6	Gaps 1
QY	1	GCCTCTACGGGGGGGGGCAAGCCTTCTCTGAGCGCTTCGCGCGTGTGTGGGCGGAGT	60	
Db	62632	GGCGGCTTACGGGGGGGGCCCAAGCCTTCTCTGACGCGCTCGCGCTGACCCGCTGGCAAGC	62693	
QY	61	GGGCTGACGGGTCACTCGATCGCTTGGGGTGTGTGGGCGGGGAGAACATGACCGGTACC	120	
Db	62692	GGCTGTCCCGGCAAGGCGATGTGCTGTGGGGGCGGTGTGGCGGGCGGGCGGAGATGGGGAGACGGC	62755	
QY	121	GAGGGCGGGGACTACTGTGGGAGCGGAGGGGCTTGGCGGCGCATGAGACCCGCAAGCGGGCGATC	180	

Db 62752 GAGGGGAAACGCTGCTGAGCCGCGGTGCTGTATAGGCGATGAGACCGCGCGCGCTG 62811  
Qy 181 GAGAGCTGCGGACACACCTTGAACGCGGGGAAACCCCTGGGTGTCTGGTGTGACCTGAC 240  
Db 62812 GCGGCACTGGGCGCGGCTTGTGTGAGAGACCTTCACTGTGACGAGGCGGACCTGGAC 62871  
Qy 241 CGGAGCGGTTCTGTGACCTGTTCAACGCGCGCGCGCGCGCGCGCTTCACTGACGACCT- 239  
Db 62872 CGGCGCGGATTCGCGGCGGCTTACCTCCGCGCGCTGATCCGCGGACCTG 62931  
Qy 300 -----CGTGGGCTCGCGCGCGGCGGAGAGACCGGTCAAGATCTCGCGCG 354  
Db 62932 ATGACGCGGAGCGCGGACCGGACCGCGCGCGCGCGCGCGCGCGGAGTGTGGAC 62991  
Qy 355 CGGCTGCGCTGACATGCGCGGAGCGGACGTCAGAGATGTGCGCGCGGCTGTGACGCG 414  
Db 62992 CGGCGGCTGACCGCGCTGCGCGCGCGCGCTGCGCGCGGACCTGTGACCTGTGCGCG 63051  
Qy 415 GAGTGGCAGCGGCTGCTGAGCGGACCGGACCGGATGTGAGCGGTGACGCTGCGCTTC 474  
Db 63052 GAGTGGCGCGGACCTGCGGACCGGCGGCGGCTGAGGCGGATGACCGGACCGGCGCTTC 63111  
Qy 475 CGTACCTGAGATTCATCTGATGACCGCGCTGACCTGCGGAGACCGGCTGCGCGCTG 534  
Db 63112 CGGACCTGCGGCTTCACTGCTGCGCGCGCGCTGCGGACCTGCGGACCGGATGCGCGAG 63171  
Qy 535 ACCGGGCTCGGCTGCGGACCGGACCGGATGTCTTGAACACCGGACGAGTGAACCGCTCAC 594  
Db 63172 ACCGGGCTACCTGCGGCGGACCGGATGTGATGACGACGACGACCGGCGCGCTGCG 63231  
Qy 595 GCGGAC 600  
Db 63232 GCGACAC 63237

RESULT 12  
US-10-760-493-22  
; Sequence 22, Application US/10760493  
; Publication No. US20050187167A1  
; GENERAL INFORMATION:  
; APPLICANT: Ecopia Biosciences Inc  
; APPLICANT: Farnet, Chris M.  
; APPLICANT: McAlpine, James B.  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Bachmann, Brian O.  
; APPLICANT: Pirae, Mahmood  
; TITLE OF INVENTION: POLYMER POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE  
; FILE REFERENCE: 3004-9US  
; CURRENT APPLICATION NUMBER: US/10/760,493  
; CURRENT FILING DATE: 2004-01-21  
; PRIOR APPLICATION NUMBER: USSN 60/441,123  
; PRIOR FILING DATE: 2003-01-21  
; PRIOR APPLICATION NUMBER: USSN 60/469,810  
; PRIOR FILING DATE: 2003-05-13  
; PRIOR APPLICATION NUMBER: USSN 60/491,516  
; PRIOR FILING DATE: 2003-08-01  
; PRIOR APPLICATION NUMBER: USSN 60/494,568  
; PRIOR FILING DATE: 2003-08-13  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 24444  
; TYPE: DNA  
; ORGANISM: Streptomyces albusensis  
US-10-760-493-22

Query Match 33.1%; Score 199; DB 9; Length 24444;  
Best Local Similarity 60.5%; Freq. No. 5, 9e-40;  
Matches 378; Conservative 0; Mismatches 220; Indels 27; Gaps 2;  
Qy 3 CTCCTACGCGGCGGACGCTTCTCTGACGCGCTTGGCGCGTGTGTGCGGAGTGG 62

Db 22559 CGCTTACGCGGCGGCGGACCACTGAGACGCGCTGCGGCGGACGCGGCGGCGGCGG 23618  
Qy 63 GCTGCGGCTGCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 119  
Db 23619 ACTGACGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23678  
Qy 120 CGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 179  
Db 23679 CGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23738  
Qy 180 CGAGGCTGCGGACCACTGAGCGGCGGAGACCGGCTGCTGCTGCTGCTGCTGCTGCTG 239  
Db 23739 CGCGCGCTGCGGACGCGGCTGAGCGGCGGAGAGAGAGGCGGTACGCTGCGGATGCTG 23798  
Qy 240 CGGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 239  
Db 23799 CTGGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23858  
Qy 300 CGGTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 335  
Db 23859 CGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23918  
Qy 336 GGAATGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 395  
Db 23919 AGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23978  
Qy 396 CGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 455  
Db 23979 CTGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24038  
Qy 456 CGAGGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 515  
Db 24039 CGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24098  
Qy 516 GAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 575  
Db 24099 CAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24158  
Qy 576 GACAGTGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db 24159 GACCGCGGCGGCACTGCGCGGCGAC 24183

RESULT 13  
US-10-156-761-412  
; Sequence 412, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMDRA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 412  
; LENGTH: 18435  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(18435)  
US-10-156-761-412



Query Match 33.0%; Score 198.4; DB 6; Length 18435;  
Best Local Similarity 60.4%; Pred. No. 8.8e-40;  
Matches 378; Conservative 0; Mismatches 221; Indels 27; Gaps 2;

Qy 1 GCGCTCTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCTGCTGCTGCGCGAGT 60  
Db 13027 GCGGCTTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCTGCTGCTGCGCGAGT 13086

Qy 61 GGGCTGCGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
Db 13087 GGGCTGCGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13146

Qy 121 GAGGCGGCGCACTA---CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 177  
Db 13147 GAGGCGGCGCACTA---CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13206

Qy 178 ATCGAGGAGCTGCGGAGCACTGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCTG 237  
Db 13207 GTGCGCGGAGCTGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCTG 13266

Qy 238 GACCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297  
Db 13267 GACCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13326

Qy 298 CTGCGTGGG---TCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333  
Db 13327 CTGCGTGGG---TCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13386

Qy 334 CAGGATCGATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393  
Db 13387 GCGCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13446

Qy 394 GTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453  
Db 13447 CTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13506

Qy 454 ATCGAGGCTGATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513  
Db 13507 GTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13566

Qy 514 CGGAACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573  
Db 13567 CGGAACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 13626

Qy 574 CGGACAGTGAACCGCTTCAACCGGCA 599  
Db 13627 CCGAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 13652

RESULT 14  
US-10-156-761-15103  
; Sequence 15103, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 15103  
; LENGTH: 100000

TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-15103

Query Match 33.0%; Score 198.4; DB 6; Length 100000;  
Best Local Similarity 60.4%; Pred. No. 6.3e-40;  
Matches 378; Conservative 0; Mismatches 221; Indels 27; Gaps 2;

Qy 1 GCGCTCTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCTGCTGCTGCGCGAGT 60  
Db 63550 GCGGCTTACGCGGCGGCAACGCGCTTCTCGACGCGCTTCCGCGCTGCTGCTGCGCGAGT 63609

Qy 61 GGGCTGCGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
Db 63610 GGGCTGCGGCTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63669

Qy 121 GAGGCGGCGCACTA---CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 177  
Db 63670 GAGGCGGCGCACTA---CTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63729

Qy 178 ATCGAGGAGCTGCGGAGCACTGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCTG 237  
Db 63730 GTGCGCGGAGCTGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCGGCGGAGCTG 63789

Qy 238 GACCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297  
Db 63790 GACCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63849

Qy 298 CTGCGTGGG---TCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 333  
Db 63850 CTGCGTGGG---TCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63909

Qy 334 CAGGATCGATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 393  
Db 63910 GCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 63969

Qy 394 GTCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453  
Db 63970 CTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 64029

Qy 454 ATCGAGGCTGATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513  
Db 64030 GTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64089

Qy 514 CGGAACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 573  
Db 64090 CGGAACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64149

Qy 574 CGGACAGTGAACCGCTTCAACCGGCA 599  
Db 64150 CCGAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 64175

RESULT 15  
US-10-156-761-411  
; Sequence 411, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02





**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:28:59 ; Search time 99.13 Seconds

(without alignments)  
10776.903 Million cell updates/sec

Title: US-10-611-442-2\_COPY\_1200\_1800

Perfect score: 1 gctcctcagcgcgcggaac.....ggacgcgcctcagcgcgacact 601

Sequence:

Scoring table: IDENTITY\_NUC

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Issued Patents NA:\*

1: /cgn2\_6/prodata/1/ina/1 COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5 COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*  
7: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*  
8: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*  
9: /cgn2\_6/prodata/1/ina/6 COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	601	100.0	47981	3	US-07-679-279-1
2	397.8	66.2	20235	2	US-07-642-734C-3
3	397.8	66.2	20235	2	US-08-439-009A-3
4	202.6	33.7	11219	2	US-07-642-734C-1
5	202.6	33.7	11219	2	US-08-439-009A-1
6	195.6	32.5	80161	3	US-09-376-987A-1
7	195.6	32.5	80161	3	US-09-376-987A-1
8	195.6	32.5	80161	3	US-09-603-207-1
9	182.4	30.3	44377	2	US-08-804-227C-7
10	182.4	30.3	44377	2	US-08-804-227C-1
11	172.6	28.7	43280	2	US-08-804-227C-1
12	172.6	28.7	43280	2	US-09-105-537-32
13	172.2	28.7	36778	3	US-09-105-537-5
14	172.2	28.7	36778	3	US-09-105-537-5
15	172.2	28.7	36778	3	US-09-105-537-5
16	172.2	28.7	36778	3	US-09-105-537-5
17	172.2	28.7	36778	3	US-09-105-537-5
18	171.6	28.6	15872	3	US-09-105-537-1
19	171.6	28.6	15872	3	US-09-105-537-1
20	171.6	28.6	15872	3	US-09-105-537-1
21	161	26.8	13842	3	US-09-105-537-30
22	160.8	26.3	50937	3	US-09-105-537-34
23	158.2	26.3	50937	3	US-09-105-537-34
24	153	25.5	77536	3	US-09-410-551B-1

25	153	25.5	77536	3	US-09-940-316B-1	Sequence 1, Appl1
26	150.4	25.0	6210	3	US-10-212-962-1	Sequence 1, Appl1
27	147	24.5	31422	3	US-09-914-286-2	Sequence 2, Appl1
28	144	24.0	53799	3	US-10-042-665A-3	Sequence 3, Appl1
29	143.8	23.9	68750	3	US-09-335-409-1	Sequence 1, Appl1
30	143.8	23.9	68750	3	US-09-568-102-1	Sequence 1, Appl1
31	143.8	23.9	68750	3	US-09-567-969-1	Sequence 1, Appl1
32	143.8	23.9	68750	3	US-09-568-486-1	Sequence 1, Appl1
33	143.8	23.9	68750	3	US-09-568-486-1	Sequence 1, Appl1
34	143.8	23.9	68750	3	US-09-568-472-1	Sequence 1, Appl1
35	143.8	23.9	68750	3	US-09-567-899-1	Sequence 1, Appl1
36	143.8	23.9	68750	3	US-10-014-717-1	Sequence 1, Appl1
37	142.2	23.7	71989	3	US-09-443-501A-2	Sequence 2, Appl1
38	142.2	23.7	71989	3	US-09-724-889A-2	Sequence 2, Appl1
39	142.2	23.7	71989	3	US-09-724-876-2	Sequence 2, Appl1
40	135.2	22.5	77536	3	US-09-410-551B-1	Sequence 1, Appl1
41	135.2	22.5	77536	3	US-09-940-316B-1	Sequence 1, Appl1
42	134.8	22.4	28958	2	US-08-258-261B-6	Sequence 6, Appl1
43	134.8	22.4	28958	2	US-08-456-837-6	Sequence 6, Appl1
44	134.8	22.4	28958	2	US-08-457-342-6	Sequence 6, Appl1
45	134.8	22.4	28958	2	US-08-457-646A-6	Sequence 6, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-679-279-1  
Sequence 1, Application US/09679279  
Patent No. 6524841  
GENERAL INFORMATION:  
APPLICANT: McDaniel, Robert  
APPLICANT: Volchegursky, Yanina  
TITLE OF INVENTION: Recombinant Megalocin Biosynthetic  
FILE REFERENCE: 300622004700  
CURRENT APPLICATION NUMBER: US/09/679, 279  
CURRENT FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/158, 305  
PRIOR FILING DATE: 1999-10-08  
PRIOR APPLICATION NUMBER: US 60/190, 024  
PRIOR FILING DATE: 2000-03-17  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 47981  
TYPE: DNA  
ORGANISM: Micromonospora megalomicea  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(144)  
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;  
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (928)...(2061)  
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;  
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (2072)...(3382)  
OTHER INFORMATION: megDII, rhodosaminyl transferase (eryCIII homolog),  
OTHER INFORMATION: TDP-megosamine glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (3462)...(4634)  
OTHER INFORMATION: megG (megY), mycarosyl acyltransferase, mycarose O-acetyltransferase;  
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (4651)...(5775)  
OTHER INFORMATION: megDII, deoxybugar transaminase (eryCI, DuxI homolog),  
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;  
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence  
NAME/KEY: CDS

LOCATION: (5822)...(6595)  
OTHER INFORMATION: megDI, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (6592)...(7197)  
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmnu hc  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (7220)...(8206)  
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmnu homolog),  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (8228)...(9220)  
OTHER INFORMATION: megBII-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (9226)...(10479)  
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (10483)...(11424)  
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (12181)...(22821)  
OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (12505)...(13470)  
OTHER INFORMATION: megAI, AT-L  
NAME/KEY: misc feature  
LOCATION: (13576)...(13791)  
OTHER INFORMATION: megAI, ACP-L  
NAME/KEY: misc feature  
LOCATION: (13849)...(15126)  
OTHER INFORMATION: megAI, KS1  
NAME/KEY: misc feature  
LOCATION: (15427)...(16476)  
OTHER INFORMATION: megAI, AT1  
NAME/KEY: misc feature  
LOCATION: (17155)...(17694)  
OTHER INFORMATION: megAI, KRI  
NAME/KEY: misc feature  
LOCATION: (17917)...(18207)  
OTHER INFORMATION: megAI, ACP1  
NAME/KEY: misc feature  
LOCATION: (18268)...(19548)  
OTHER INFORMATION: megAI, KS2  
NAME/KEY: misc feature  
LOCATION: (19876)...(20910)  
OTHER INFORMATION: megAI, AT2  
NAME/KEY: misc feature  
LOCATION: (21517)...(22053)  
OTHER INFORMATION: megAI, KR2  
NAME/KEY: misc feature  
LOCATION: (22318)...(22575)  
OTHER INFORMATION: megAI, ACP2  
NAME/KEY: CDS  
LOCATION: (22867)...(33555)  
OTHER INFORMATION: megAI, SEQ ID NO: 14= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (32957)...(4237)  
OTHER INFORMATION: megAI, KS3  
NAME/KEY: misc feature  
LOCATION: (24514)...(25581)  
OTHER INFORMATION: megAI, AT3  
NAME/KEY: misc feature  
LOCATION: (26230)...(26733)  
OTHER INFORMATION: megAI, KR3 (inactive)  
NAME/KEY: misc feature  
LOCATION: (26958)...(27258)

OTHER INFORMATION: megAI, ACP3  
NAME/KEY: misc feature  
LOCATION: (27353)...(28590)  
OTHER INFORMATION: megAI, KS4  
NAME/KEY: misc feature  
LOCATION: (28977)...(29931)  
OTHER INFORMATION: megAI, AT4  
NAME/KEY: misc feature  
LOCATION: (29953)...(30477)  
OTHER INFORMATION: megAI, DH4  
NAME/KEY: misc feature  
LOCATION: (31396)...(32244)  
OTHER INFORMATION: megAI, ER4  
NAME/KEY: misc feature  
LOCATION: (32257)...(32799)  
OTHER INFORMATION: megAI, KR4  
NAME/KEY: misc feature  
LOCATION: (33052)...(33312)  
OTHER INFORMATION: megAI, ACP4  
NAME/KEY: CDS  
LOCATION: (33666)...(43271)  
OTHER INFORMATION: megAI, SEQ ID NO: 15= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (33780)...(35027)  
OTHER INFORMATION: megAI, KS5  
NAME/KEY: misc feature  
LOCATION: (35385)...(36419)  
OTHER INFORMATION: megAI, AT5  
NAME/KEY: misc feature  
LOCATION: (37068)...(37604)  
OTHER INFORMATION: megAI, KR5  
NAME/KEY: misc feature  
LOCATION: (37860)...(38120)  
OTHER INFORMATION: megAI, ACP5  
NAME/KEY: misc feature  
LOCATION: (38187)...(39470)  
OTHER INFORMATION: megAI, KS6  
NAME/KEY: misc feature  
LOCATION: (39755)...(40811)  
OTHER INFORMATION: megAI, AT6  
NAME/KEY: misc feature  
LOCATION: (41406)...(41936)  
OTHER INFORMATION: megAI, KR6  
NAME/KEY: misc feature  
LOCATION: (42168)...(42425)  
OTHER INFORMATION: megAI, ACP6  
NAME/KEY: misc feature  
LOCATION: (42585)...(43271)  
OTHER INFORMATION: megAI, TE  
NAME/KEY: CDS  
LOCATION: (43268)...(44344)  
OTHER INFORMATION: megII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;  
OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (44355)...(45623)  
OTHER INFORMATION: megCII, desosaminyl transferase, desosamine glycosyltransferase  
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;  
OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (45620)...(46591)  
OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,  
OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxyglucose 2,3 dehydratase;  
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;  
OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (46660)...(47403)  
OTHER INFORMATION: megH, TEII, SEQ ID NO: 19= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (47411)...(47980)  
OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequ  
us-09-679-279-1

Query Match 100.0%; Score 601, DB 3, Length 47981;  
Best Local Similarity 100.0%; Pred. No. 2,5e-104, Indels 0, Gaps 0;  
Matches 601, Conservative 0, Mismatches 0;

QY	1	GCCTCTACACGCGGCGGCAACGCTTCTCTGACGCTTGGCCGCTGTGGGCGCACT	60
Db	37506	GCCTCTACACGCGGCGGCAACGCTTCTCTGACGCTTGGCCGCTGTGGGCGCACT	37565
QY	61	GCGCTGCGCGGTCACTTCGATCGCTTGGGGGTCTGTGGGCGGAGCAATGCGCGTACC	120
Db	37566	GCGCTGCGCGGTCACTTCGATCGCTTGGGGGTCTGTGGGCGGAGCAATGCGCGTACC	37625
QY	121	GAGGCGCGCGCATCACTCGGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG	180
Db	37626	GAGGCGCGCGCATCACTCGGCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGG	37685
QY	181	GAGGAGCTGCGGACCACTTCGACGCGGCGGACCGCGGCGGCGGCGGCGGCGG	240
Db	37686	GAGGAGCTGCGGACCACTTCGACGCGGCGGACCGCGGCGGCGGCGGCGGCGG	37745
QY	241	CGGAGCGCGGTTCTGTGAACTGTTCAACCGCGCGCGCGCGCGCGCGCTTTCGACGAACTC	300
Db	37746	CGGAGCGCGGTTCTGTGAACTGTTCAACCGCGCGCGCGCGCGCGCGCTTTCGACGAACTC	37805
QY	301	GCTGGGCGTCCG	360
Db	37806	GCTGGGCGTCCG	37865
QY	361	GCGTCGATGCGCGAGCGCGGACGTCAGACATGTCGCGCGCGCGCGCGCGCGCGCG	420
Db	37866	GCGTCGATGCGCGAGCGCGGACGTCAGACATGTCGCGCGCGCGCGCGCGCGCG	37925
QY	421	GCAAGCGGCTGTGGGCGCAACGCGCAACGCGCGGATGAGCGTACGTCGCTTCGATGAC	480
Db	37926	GCAAGCGGCTGTGGGCGCAACGCGCAACGCGCGGATGAGCGTACGTCGCTTCGATGAC	37985
QY	481	CTGGGATTTCCATCTTCATGACACCGCGCTGACACTTCGCGGACACCGGCTCGCGCGCGG	540
Db	37986	CTGGGATTTCCATCTTCATGACACCGCGCTGACACTTCGCGGACACCGGCTCGCGCGCGG	38045
QY	541	GTCGCGGTGCGCAACGACATCGTCTTCGACCAACCGGACGATGACGCGCTCACCGCGCAC	600
Db	38046	GTCGCGGTGCGCAACGACATCGTCTTCGACCAACCGGACGATGACGCGCTCACCGCGCAC	38105
QY	601	T 601	
Db	38106	T 38106	

RESULT 2  
US-07-642-734C-3  
: Sequence 3, Application US/07642734C  
: Patent No. 5824513  
: GENERAL INFORMATION:  
: APPLICANT: Katz, L  
: APPLICANT: Donald, S  
: APPLICANT: McAlpine, J B  
: TITLE OF INVENTION: Recombinant DNA Method for Producing  
: TITLE OF INVENTION: Erythromycin Analogs  
: NUMBER OF SEQUENCES: 27  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Edward H. Gorman  
: STREET: Abbott Laboratories D377/AbPd-2 One Abbott  
: STREET: Park Rd  
: CITY: Abbott Park  
: STATE: IL  
: COUNTRY: US  
: ZIP: 60064-3500  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/07/642,734C  
: FILING DATE: 17-JAN-91

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dancigers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start=19
OTHER INFORMATION: /functions="gene eryA"
OTHER INFORMATION: /product="eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..3470
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..22670
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..33921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: dehydroatase and enoylreductase domains m"
FEATURE:

```

```

NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf1 encoding modules 5 & 6
6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10773..15165
OTHER INFORMATION: /function= "approximate span of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12374
OTHER INFORMATION: /function= "approximate span of module 5"
OTHER INFORMATION: beta-ketoacylACPythnase domain of modul
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of module 5"
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14052..14610
OTHER INFORMATION: /function= "approximate span of module 5"
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of module 5"
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: beta-ketoacylACPythnase domain of modul
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19145..19398
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of module 6"
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

```

```

Query March 66.2%; Score 397.8; DB 2; Length 20235;
Best Local Similarity 78.9%; Pred. No. 3,3e-66;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1 GCTCTCTACGCGGCGGCGCAACGCTTCTCTGACGCGCTTGCGCCGCTGTGTGCGCGCACT 60
DB 14500 GCTCTCTACGCGGCGGCGCAACGCTTCTCTGACGCGCTTGCGCCGCTGTGTGCGCGCACT 14559
QY 61 GGGCTGCGCGGTGACCTTCGATTCGCTGTGGGTCTGTGGGCGGCGAGAACATGGCCGGTACC 120
DB 14560 GGGCTGCGCGGTGACCTTCGATTCGCTGTGGGTCTGTGGGCGGCGAGAACATGGCCGGTACC 14619
QY 121 GAGGCGCGCGCACTACCTGTGCGAGCGAGCGGCGCTGCGCGCATGAGACCCGAGCGGCGATC 180
DB 14820 GAGGCGCGCGCACTACCTGTGCGAGCGAGCGGCGCTGCGCGCATGAGACCCGAGCGGCGATC 14679
QY 181 GAGGAGCTGCGAGCAACCTTGACGCGCGGAGACCTGTGGGTGTCCGTGTGTGACCTTGAC 240
DB 14680 GAGGAGCTGCGAGCAACCTTGACGCGCGGAGACCTGTGGGTGTCCGTGTGTGACCTTGAC 14739
QY 241 CGGAGAGCGTTTCTGTGCACTGTTCAACCGCGCGCGCGCGCGCGCGCGCGCTTTGACGACATC 300
DB 14740 CGGAGAGCGTTTCTGTGCACTGTTCAACCGCGCGCGCGCGCGCGCGCGCGCTTTGACGACATC 14799
QY 301 GGTGGGCTCCGCGCGCGGCGCGAGAGACCGGTCAGAGATCGGATCTGCCCGCGCGCTG 360
DB 14800 GGTGGGCTCCGCGCGCGGCGCGAGAGACCGGTCAGAGATCGGATCTGCCCGCGCGCTG 14859
QY 361 GCGTCGATGCGCGAGCGCGCAACGTCACAGACATGTGCGCCGCGCTGTCGAGCCGAGGTC 420
DB 14860 GCGTCGATGCGCGAGCGCGCAACGTCACAGACATGTGCGCCGCGCTGTCGAGCCGAGGTC 14919
QY 421 GCAAGCGGTGCTGGGCGCGAGCGCGGACGCGGATGTGAGCGTGAAGCTTCCCTTCCGTGAC 480
DB 14920 GCAAGCGGTGCTGGGCGCGAGCGCGGACGCGGATGTGAGCGTGAAGCTTCCCTTCCGTGAC 14979
QY 481 CTGGGATTTGCACTTCATGACCGCGCTGACCTTGCGGACCGGCTGCGGCGGTCGACCGG 540
DB 14980 CTGGGATTTGCACTTCATGACCGCGCTGACCTTGCGGACCGGCTGCGGCGGTCGACCGG 15039
QY 541 GTCCGAGTGGCGACGACCATGCTTTCGACCAACCGGACAGATGAGCCGCTCACCGCGCAC 600
DB 15040 GTCCGAGTGGCGACGACCATGCTTTCGACCAACCGGACAGATGAGCCGCTCACCGCGCAC 15099
QY 601 T 601
DB 15100 T 15100

RESULT 3
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polypeptides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Steven F. Weinlock
; STREET: Abbott Laboratories D377/ApEd-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

1 CURRENT APPLICATION DATA:  
2 APPLICATION NUMBER: US/08/439,009A  
3 FILING DATE: 11-MAY-1995  
4 CLASSIFICATION: 435  
5 ATTORNEY/AGENT INFORMATION:  
6 NAME: Casuto, Dianne  
7 REGISTRATION NUMBER: 40,943  
8 REFERENCE/DOCKET NUMBER: 4952.US.D1  
9 TELECOMMUNICATION INFORMATION:  
10 TELEPHONE: 847-938-3137  
11 INFORMATION FOR SEQ ID NO: 3:  
12 SEQUENCE CHARACTERISTICS:  
13 LENGTH: 20235 base pairs  
14 TYPE: nucleic acid  
15 STRANDEDNESS: double  
16 TOPOLOGY: unknown  
17 MOLECULE TYPE: DNA (genomic)  
18 HYPOTHEICAL: NO  
19 ANTI-SENSE: NO  
20 ORIGINAL SOURCE:  
21 ORGANISM: Saccharopolyspora erythraea  
22 STRAIN: NRRL 238  
23  
24 FEATURE:  
25 NAME/KEY: CDS  
26 LOCATION: 19..10722  
27 OTHER INFORMATION: /codon\_start= 19  
28 OTHER INFORMATION: /function= "gene eryA"  
29 OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
30 OTHER INFORMATION: 6-deoxyerythronolide B"  
31 FEATURE:  
32 NAME/KEY: misc\_feature  
33 LOCATION: 19..4470  
34 OTHER INFORMATION: /function= "approximate span of  
35 OTHER INFORMATION: module 3"  
36 FEATURE:  
37 NAME/KEY: misc\_feature  
38 LOCATION: 97..1482  
39 OTHER INFORMATION: /function= "approximate span of  
40 OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
41 FEATURE:  
42 NAME/KEY: misc\_feature  
43 LOCATION: 1693..2670  
44 OTHER INFORMATION: /function= "approximate span of  
45 OTHER INFORMATION: acyltransferase domain module 3"  
46 FEATURE:  
47 NAME/KEY: misc\_feature  
48 LOCATION: 3406..3921  
49 OTHER INFORMATION: /function= "approximate span of  
50 OTHER INFORMATION: beta-ketoreductase domain of module 3"  
51 FEATURE:  
52 NAME/KEY: misc\_feature  
53 LOCATION: 4171..4428  
54 OTHER INFORMATION: /function= "approximate span of  
55 OTHER INFORMATION: acyl carrier domain of module 3"  
56 FEATURE:  
57 NAME/KEY: misc\_feature  
58 LOCATION: 4471..10722  
59 OTHER INFORMATION: /function= "approximate span of  
60 OTHER INFORMATION: module 4"  
61 FEATURE:  
62 NAME/KEY: misc\_feature  
63 LOCATION: 4471..5847  
64 OTHER INFORMATION: /function= "approximate span of  
65 OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"  
66 FEATURE:  
67 NAME/KEY: misc\_feature  
68 LOCATION: 6054..7026  
69 OTHER INFORMATION: /function= "approximate span of  
70 OTHER INFORMATION: acyltransferase domain of module 4"  
71 FEATURE:  
72 NAME/KEY: misc\_feature  
73 LOCATION: 7165..9216  
74 OTHER INFORMATION: /function= "approximate span of

1 OTHER INFORMATION: dehydratase and enoylreductase domains m"  
2 FEATURE:  
3 NAME/KEY: misc\_feature  
4 LOCATION: 9433..9984  
5 OTHER INFORMATION: /function= "approximate span  
6 OTHER INFORMATION: beta-ketoreductase of module 4"  
7 FEATURE:  
8 NAME/KEY: misc\_feature  
9 LOCATION: 10225..10483  
10 OTHER INFORMATION: /function= "approximate span of  
11 OTHER INFORMATION: acyl carrier domain of module 4"  
12 FEATURE:  
13 NAME/KEY: CDS  
14 LOCATION: 10723..20235  
15 OTHER INFORMATION: /codon\_start= 10723  
16 OTHER INFORMATION: /function= "gene =eryA"  
17 OTHER INFORMATION: /product= "orf13 encoding modules 5 & 6  
18 OTHER INFORMATION: 6-deoxyerythronolide B formatio"  
19 FEATURE:  
20 NAME/KEY: misc\_feature  
21 LOCATION: 10723..15165  
22 OTHER INFORMATION: /function= "approximate span of  
23 OTHER INFORMATION: module 5"  
24 FEATURE:  
25 NAME/KEY: misc\_feature  
26 LOCATION: 10831..12174  
27 OTHER INFORMATION: /function= "approximate span of  
28 OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
29 FEATURE:  
30 NAME/KEY: misc\_feature  
31 LOCATION: 12375..13350  
32 OTHER INFORMATION: /function= "approximate span of  
33 OTHER INFORMATION: acyltransferase domain of module 5"  
34 FEATURE:  
35 NAME/KEY: misc\_feature  
36 LOCATION: 14062..14610  
37 OTHER INFORMATION: /function= "approximate span of  
38 OTHER INFORMATION: beta-ketoreductase of module 5"  
39 FEATURE:  
40 NAME/KEY: misc\_feature  
41 LOCATION: 14857..15114  
42 OTHER INFORMATION: /function= "approximate span of  
43 OTHER INFORMATION: acyl carrier domain of module 5"  
44 FEATURE:  
45 NAME/KEY: misc\_feature  
46 LOCATION: 15166..20235  
47 OTHER INFORMATION: /function= "approximate span of  
48 OTHER INFORMATION: module 6"  
49 FEATURE:  
50 NAME/KEY: misc\_feature  
51 LOCATION: 15172..16569  
52 OTHER INFORMATION: /function= "approximate span of  
53 OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
54 FEATURE:  
55 NAME/KEY: misc\_feature  
56 LOCATION: 16768..17721  
57 OTHER INFORMATION: /function= "approximate span of  
58 OTHER INFORMATION: acyltransferase domain of module 6"  
59 FEATURE:  
60 NAME/KEY: misc\_feature  
61 LOCATION: 18375..18921  
62 OTHER INFORMATION: /function= "approximate span of  
63 OTHER INFORMATION: beta-ketoreductase domain of module 6"  
64 FEATURE:  
65 NAME/KEY: misc\_feature  
66 LOCATION: 19145..19398  
67 OTHER INFORMATION: /function= "approximate span of  
68 OTHER INFORMATION: acyl carrier domain of module 6"  
69 FEATURE:  
70 NAME/KEY: misc\_feature  
71 LOCATION: 19492..20235  
72 OTHER INFORMATION: /function= "approximate span of  
73 OTHER INFORMATION: thioesterase domain of module 6"



NAME/KEY: misc.feature  
LOCATION: 6678..8066  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 2"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 8262..9305  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 2"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 9906..10454  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 2"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 10707..10964  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 2"  
US-07-642-734C-1

Query Match 33.7%; Score 202.6; DB 2; Length 11219;  
Best Local Similarity 60.9%; Pred. No. 1,4e-29;  
Matches 366; Conservative 0; Mismatches 229; Indels 6; Gaps 2;

QY 1 GCCTCTACGCGGCGGCAAGCCTTCTGAGCCCTTCCGCGCTGCTGCGGCGAGT 60  
DB 10344 GGGCGCTACGCGCGGCAAGCCTTCTGAGCCCTTCCGCGCGGCGGCGGCG 10403  
QY 61 GGGCTGCGGCTGACCTGATGCGCTGCGGCTGCTGCGGCGGCGGCAAGATGCGGTAAC 120  
DB 10404 GGGCTGCGGCGCAAGCTGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10463  
QY 121 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 177  
DB 10464 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10523  
QY 178 ATGAGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 237  
DB 10524 CTGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10583  
QY 238 GACCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297  
DB 10584 GACTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10643  
QY 298 CTGCGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 354  
DB 10644 ATGCTACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 10703  
QY 355 CGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 414  
DB 10704 CGGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10763  
QY 415 GAGCTGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 474  
DB 10764 AGCAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10823  
QY 475 CGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 534  
DB 10824 AAGGAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10883  
QY 535 ACCGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 594  
DB 10884 ACCGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10943  
QY 595 G 595  
DB 10944 G 10944

RESULT 5  
US-08-439-009A-1  
; Sequence 1, Application us/08439009A

Patent No. 6004787  
GENERAL INFORMATION:  
APPLICANT: Donadio, S  
APPLICANT: Katz, L  
APPLICANT: McAlpine, J B  
TITLE OF INVENTION: Method of Directing Biosynthesis of  
TITLE OF INVENTION: Specific Polypeptides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Steven F. Weinstein  
STREET: Abbott Laboratories D377/Aped-2 One Abbott  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/439,009A  
FILING DATE: 11-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Casato, Diane  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 2338  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 744..6659  
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF  
OTHER INFORMATION: MODULE 1"  
OTHER INFORMATION: /label= FUNCTION  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 744..11219  
OTHER INFORMATION: /function= "gene= "eryA"  
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 744..1868  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1998..2198  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 2250..3626  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 3831..4811





Dh 39481 GGTCTCCGCGCAACTCGATCTCTG968GAGTTG96CCG9CG9CATG9CGAAGG 39540  
Qy 121 GAGGCGGCGACTACCTGCGCAGCCAGGCGCTGCGCGCAATGGAACCGCGG9GATC 180  
Dh 39541 GCGGCGGCGGAACTTGGCGGCAACCGG9ATTCGTCGATGCGG9GCGTGGCCATC 39600  
Qy 181 GAGAGCTGCGGCAACCTTGAAGCGCGG9ACCCGTG9GTGTCGTGTCGACCTG9AC 240  
Dh 39601 CTGGCTCTGCAAGAAAGTACTTGACAGAGATGAGACGTGGGTGTGATGCTGATGTGAC 39660  
Qy 241 CCGGAGCGGTCTCTCAACTGTTCACTGCGCGCGCGCGCGCGCCCTCTTGAAGAACTC 300  
Dh 39661 TGGAGCCGATTCCTTCCACGTTTCGCGCATGCGCGCAACCGGTTGTTCAGAGATG 39720  
Qy 301 GGTGGGATCCGCGCGCGGCGGCGGAGAACCGGTCAAGAA-----TCGATCTC 348  
Dh 39721 CCGGCGGCGGAAAGGCGATGCGCGGATG9GCGCGGAGAACAGGCGGCTGCGGTTT 39780  
Qy 349 GCGCGGCGGCTGCGCTGATGTCGAGAGCGGAAAGTCAAGAGATGTCGCGCGCTGATC 408  
Dh 39781 GCGCGCAATCTGCGGAGGTCGCGGAGGCGCAAGCGCAAGCACTGTGTGATCTGATC 39840  
Qy 409 CGAGCGAGGTGCGAGCGGTCTG9GCGCAAGCGCAAGCGGATGATCGAGCTGACGTC 468  
Dh 39841 TGGCGCCAGGTGCGAACCGGTCTG9GCGCAAGCGGATGCGAGGAGATCCAGCCGAGCG 39900  
Qy 469 GCGTTCGATGACCTGAGATTCGATCTCATGACCGCGCTGATCCTTGGGAAACCGGCTCGG 528  
Dh 39901 GCGTTCGCGCGCTGCGGTTCGATCTCTCATGCGGTGATCTTGGCAATCTTTGATC 39960  
Qy 529 GCGGTGACCGG9GTCCG9GTGCGCACAGCATGCTTTCGACCAAGCGGAGTGAAGCGG 588  
Dh 39961 ACCGCGACCGG9GTGCGCTGCGGACCAACCGTCTTGACCTACCGGATCCGCGCGCC 40020  
Qy 589 CTCACCGGCGAC 600  
Dh 40021 TTGGCGGCTCAC 40032

## RESULT 7

US-09-370-700-1  
; Sequence 1, Application US/09370700  
; Patent No. 6274350  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 Div1  
; CURRENT APPLICATION NUMBER: US/09/370,700  
; EARLIER FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/36987  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-1

Query Match 32.5%; Score 195.6; DB 3; Length 80161;  
Best Local Similarity 59.0%; Pred. No. 2.8e-28;  
Matches 361; Conservative 0; Mismatches 239; Indels 12; Gaps 1;  
Qy 1 GCCTCTACGCGCGGCAAGCGCTTCTCGACGCTTGCCTGTCGTGTGTCGCGCACT 60  
Dh 39421 GCGGCTTACGCGCGGCAAGCGCATTTCTTGACACACTGCGGAAACAGCGCGAGCAGCC 39480

Qy 61 GGGCTCCG9TACACTCGATTCGCTTGG9GTCTGTG9GCGCGGAGAAATGCGCGATACC 120  
Dh 39481 GGTCTCCGCGCAACTCGATCTCTTGG9GCACTTGG9GCGCGGCAATGCGCGAGCGC 39540  
Qy 121 GAGGCGGCGACTACCTGCGCAGCCAGGCGCTGCGCGCAATGGAACCGGAGCG9GATC 180  
Dh 39541 GCGGCGGCGGAACTTGGCGGCAACCGG9ATTCGTCGATGCGG9GCGGTGGCCATC 39600  
Qy 181 GAGAGCTGCGGCAACCTTGAAGCGCGG9ACCCGTG9GTGTCGTGTCGACCTG9AC 240  
Dh 39601 CTGGCTCTGCAAGAAAGTACTTGACAGAGATGAGACGTGGGTGTGATGCTGATGTGAC 39660  
Qy 241 CCGGAGCGGTCTCTCAACTGTTCACTGCGCGCGCGCGCGCGCCCTCTTGAAGAACTC 300  
Dh 39661 TGGAGCCGATTCCTTCCACGTTTCGCGCATGCGCGCAACCGGTTGTTCAGAGATG 39720  
Qy 301 GGTGGGATCCGCGCGGCGGCGGAGAACCGGTCAAGAA-----TCGATCTC 348  
Dh 39721 CCGGCGGCGGAAAGGCGATGCGCGGATG9GCGCGGAGAACAGGCGGCTGCGGTTT 39780  
Qy 349 GCGCGGCGGCTGCGCTGATGTCGAGAGCGGAAAGTCAAGAGATGTCGCGCGCTGATC 408  
Dh 39781 GCGCGCAATCTGCGGAGGTCGCGGAGGCGCAAGCGCAAGCACTGTGTGATCTGATC 39840  
Qy 409 CGAGCGAGGTGCGAGCGGTCTG9GCGCAAGCGCGCAAGCGGATGATCGAGCTGACGTC 468  
Dh 39841 TGGCGCCAGGTGCGAACCGGTCTG9GCGCAAGCGGATGCGAGGAGATCCAGCCGAGCG 39900  
Qy 469 GCGTTCGATGACCTGAGATTCGATCTCATGACCGCGCTGATCCTTGGGAAACCGGCTCGG 528  
Dh 39901 GCGTTCGCGCGCTGCGGTTCGATCTCTCATGCGGTGATCTTGGCAATCTTTGATC 39960  
Qy 529 GCGGTGACCGG9GTCCG9GTGCGCACAGCATGCTTTCGACCAAGCGGAGTGAAGCGG 588  
Dh 39961 ACCGCGACCGG9GTGCGCTGCGGACCAACCGTCTTGACCTACCGGATCCGCGCGCC 40020  
Qy 589 CTCACCGGCGAC 600  
Dh 40021 TTGGCGGCTCAC 40032

## RESULT 8

US-09-603-207-1  
; Sequence 1, Application US/09603207B  
; Patent No. 6521406  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H  
; APPLICANT: Broughton, Mary C  
; APPLICANT: Crawford, Kathryn P  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patti J  
; APPLICANT: Turner, Jan R  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489 Div1  
; CURRENT APPLICATION NUMBER: US/09/603,207B  
; EARLIER FILING DATE: 2000-06-23  
; EARLIER APPLICATION NUMBER: 09/370,700  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 80161  
; TYPE: DNA  
; ORGANISM: Saccharopolyspora spinosa  
US-09-603-207-1

Query Match 32.5%; Score 195.6; DB 3; Length 80161;  
Best Local Similarity 59.0%; Pred. No. 2.8e-28;  
Matches 361; Conservative 0; Mismatches 239; Indels 12; Gaps 1;  
Qy 1 GCCTCTACGCGCGGCAAGCGCTTCTCGACGCTTGCCTGTCGTGTGTCGCGCACT 60  
Dh 39421 GCGGCTTACGCGCGGCAAGCGCATTTCTTGACACACTGCGGAAACAGCGCGAGCAGCC 39480

Db 39421 GGGGCTTAGGGGCGGGGGAAGCATCTTGAGACACTGCGCGAACAAGCCGAGCAGC 39480  
QY 61 GGGGCTGCGGTCACTCTGATCGCTTGAGGCTCTGTGGCCCGGGGAGAACTAGGCGGTACC 120  
Db 39481 GGTCTGCGGGCAACTCTGATCTCTTGAGGAGTGTGGCCCGCGCGCATGGCGGAGCGC 39540  
QY 121 GAGGGGCGGAGCACTACCTGCGGAGCAAGGGCTGTGGCGGCATAGAACCCGCGAGGGGCGATC 180  
Db 39541 GCGGCGGGGCAACAACCTGCGGAGCAAGGGGATAGTCCATGCGCGGCGGTGGCGCATC 39600  
QY 181 GAGAGCTGCGGAGCAACCTGAGAGCGCGGGAGCCCGTGGTGTGGTGGTGAACCTTGAC 240  
Db 39601 CTGGCTTGAGAGGAGTACTTGAACCAAGATGAGAGTGGCGGTGATCGCTGATGTGAC 39660  
QY 241 CGGAGCGGTTCTGTGAACTGTTCAACCGCGCGCGCGCGCGCGCGCTTCTTGAGAACTC 300  
Db 39661 TGGGACCGATTCTGTTCCCAAGTTGCGCGGCACTGCGCAACCCGGTTGTTGAGAAAGTG 39720  
QY 301 GGTGGGGTCCGCGCGCGGGGCGAGAGAACCGGTGAGAA-----TGGATCTC 348  
Db 39721 CCGGCGGAGAGAAAGGCGATGCCGCGAATGGCGCGGAGAACCAAGCGCGCTGCGCTTC 39780  
QY 349 GCCCGCGCGCTGCGGTGATGTCGAGAGCGCGAGCGCAAGTCAAGAGCATGTGCGCGCTGATC 408  
Db 39781 GCCCGCAATCTCGCGAGCTGCGCGAGAGCCCAAGCGCAAGTGTGATGTGATG 39840  
QY 409 CGAGCCGAGTGGCGAGCGGTCTGGGCGACCGGACCGCGAGCGTGAATGAGCGTGAAGTC 468  
Db 39841 TGCCTCCGAGTGGGCAACGTGCTCGGGGACCGGAGTGGCGAGAGTCCAGCCCGGAGCGG 39900  
QY 469 GCCTTCGGTGAAGTGGGATGATGATCCATGACCGCGCGGTGATGAGCGGAAACCGGCTGCG 528  
Db 39901 GCGTTCCGCGCGCTCGGGTGAATCTCTTCAAGCGGTGATCTGCGGATCTGTTGAC 39960  
QY 529 GCGGTGACCGGAGTCCGGGTGCGGACGACATGCTTGAACAACCGCGAGTGAACCGC 588  
Db 39961 ACCGCGACCGGAGTGGCGCTGCGCGACCAACCGCTTGAATGACCGGAGTCCGCGC 40020  
QY 589 CTCACCGCGAC 600  
Db 40021 TTGCGCGCTCAC 40032

RESULT 9  
US-08-804-227C-7  
Sequence 7, Application US/08804227C  
Patent No. 587691  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuester, Stuart A.  
APPLICANT: Kuester, Paul R., Jr.  
APPLICANT: Sulton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSER: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII (DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784

REFERENCE/DOCKET NUMBER: X-8231  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ. ID NO. 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

Query Match 30.3%; Score 182.4; DB 2; Length 4437;  
Best Local Similarity 60.1%; Pred. No. 8.3e-26;  
Matches 363; Conservative 0; Mismatches 226; Indels 15; Gaps 3;

QY 7 TACGCGCGGAGCAAGCGCTTCTGACAGCGCTTCCGCGGTGATGATGCGGAGCGAGTGGCTG 66  
Db 35220 TATCGCGCGGAGCAATGCGCATCTGAGTGGCGGAGGCGTGTGTCGAGAGGCGGT 35279  
QY 67 CCGGTCACTTGATGCGCTTGGGCTGTGTGGCGCGGAGCAACATGAGCGGTATCGGAGGCG 126  
Db 35280 CCGGAGACTTCCGCGCTGCGGCGCTGTGGGCGCGGAGGAGCATGAGGAGGAGGAGC 35339  
QY 127 GGC---GACTACCTGCGGAGCGAGGCGCTGCGGAGTGAACCGCGAGCGGAGTGAAG 183  
Db 35340 GTCAAGGAGTCTTACGCGGAGCGGCGCTGCGGCGGAGCGGAGTGAAGTGAAG 35399  
QY 184 GAGCTGCGGAGCAACCTGAGAGCGCGGAGGAGCCGCTGGGTGTGATGATGATGATGATG 243  
Db 35400 GACTTGAACAGGAGTGAACGAGGAGGAGCAACCTGCGTCAAGTGGCGCATGATGATG 35459  
QY 244 GAGCGGTTCTGGAATGTTTCAACGCGCGCGCGCGCGCGCGCTTCTTGAACGATGATG 303  
Db 35460 GAAACATTCGTCACCGGAGTTCACCGCTTACCGGCGGAGCGGAGTGAATCTGCAATCCC 35519  
QY 304 GGGTTCGCGCGC-----CGGGCGGAGGAGGAGGAGTGAAGAACTGATCTGCGCGGCG 357  
Db 35520 CAGGTTCGCGGCTTTCGCGAGCGCGGAGCAACCTGAGAGCTGCGGAGGAGTGGCGG 35579  
QY 358 CT-----GCGGTGATGCGCGGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 411  
Db 35580 GCGGTGAGCGCGCGCTTCAACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 35639  
QY 412 GCGGAGTGGAGCGGTGCTGAGGCGAGCGGAGCGCGGAGGAGTGAAGTGAAGTGAAGTGA 471  
Db 35640 ACGGTGGCGGAGGAGTCTGAGTCAAGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAG 35699  
QY 472 TTCTGTAACCTGGATTTGATCTTCAATGACCGGCGGAGGAGGAGGAGGAGGAGGAGG 531  
Db 35700 TTCTGGAACCTGCGCTTGAATCTGAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 35759  
QY 532 GAGACCGGAGTCCGGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 591  
Db 35760 GCGACGGAAGTGTGATCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 35819  
QY 592 ACCG 595

Db 35820 GCGC 35823

## RESULT 10

US-08-804-198-1

Sequence 1, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burett, Stanley G.  
APPLICANT: Kumboss, Stuart A.  
APPLICANT: Rao, Nagaraja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Roelck, Paul R., Jr.  
TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-198-1

Query Match 30.3%; Score 182.4; DB 2; Length 44377;  
Best Local Similarity 60.1%; Prid. No. 8.3e-26;  
Matches 363; Conservative 0; Mismatches 226; Indels 15; Gaps 3;

Db 35220 TATGCGCGCGCAATCGATCGATGCGGCGGAGCGGTGTGTGCGCGCGGCGGT 35279  
QY 7 TACGCGCGCGCAACGCTTCTCGAGCGCTTCCCGCTGCGGTGCGGCGAGTGGCGTG 66  
CGGTAACCTTCATCGCTGCGGTCTGTGCGCGCGCGCAACAATGCGCGGTACCGAGGCG 126  
35280 CCGCGACCTTCCTCGCTGCGGCGCTGTGCGGCGCGGAGCGCATGGAACCGAGAGGCG 35339

QY 127 GGC---GACTACTCGCGACGCCAGGCGCTGCGCGCCATGGAACCGGAGCGGCGATGAG 183  
Db 35340 GTTACGGAAGTTTCTACGCCAGCGCGGCTTCCGCCATGCGGCGCCAGTGGGCGATGAG 35399  
QY 184 GACTCGCGACCACTTGAACCGCGGAGACCGGTGAGTGTGCGGTGAGCACTGAGCCG 243  
Db 35400 GCACTGCAACCGCACTGAACGAGGCGGACACCTGCGTCACTGCGCGCATGCACTGG 35459  
QY 244 GAGCGGTTGTCGAATCTGTTACCGCGCGCGCGCGCGCGCGCGCTCTTTCAGCACTGAGT 303  
Db 35460 GAACACTGTGTCACCGGTTTCAACCGCTTACCGGCGCGCGCGCGGTGATCTTCCGACATCC 35519  
QY 304 GGGGTCGCGCGC-----CGGGCGCGAGAGACCGGTCAAGAAATCGGATCTCGCCGCGCG 357  
Db 35520 CAGTCCGCGCGCTTGGCGGACCGCCGAAACCACTGTGAGCGCTTGAAGGACTGCGCGG 35579  
QY 358 CT-----GGCTGATGCGCGAGCGCGGAGCGGACGTCAAGACATGTCGCGGCTGATCGGA 411  
Db 35580 CGGTCGACGCGCGCTTCAACCGCGCGGAGCGGACCAAGTCTGTGTCGACTGATCGG 35639  
QY 412 GCGGAGTGGCAGCGGTGCTGGGCGGACCGGACCGGACCGGTATCGAGCTGACGTGCC 471  
Db 35640 ACGGTGCGCGCGAGGTCTTCTGCTCAACGAGATCGGCGGATCGGCGGACGAGCGGCG 35699  
QY 472 TTCGATGACCTGGAATTCGACTTCATGACCGCGGTGACCTGCGGAAACCGGCTCGGCGG 531  
Db 35700 TTCGGAACCTCGGCTTCACTGCTGCGCGCGGTGCGGATGCGGCGGCGGCGGCGGAG 35759  
QY 532 GTGACCGGCGGTGCGGCGGCGGACGACCATGCTTTCGACCAACCGGACAGTGAACCGCTC 591  
Db 35760 GCGACCGGACTGTGATGCGCGCGGACGCGTCACTTTCGACCAACCGGACAGCGGCTC 35819  
QY 592 ACCG 595  
Db 35820 GCGC 35823

## RESULT 11

US-08-804-227C-1

Sequence 1, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kumboss, Stuart A.  
APPLICANT: Roelck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43280 base pairs

RESULT 12  
US-09-105-537-32  
; Sequence 32, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.

```

RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-07-26
; NUMBER OF SEQ ID NOS: 43

```

SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-5

Query Match 28.7%; Score 172.2; DB 3; Length 36778;  
Best Local Similarity 57.3%; Pred. No. 6.8e-24;  
Matches 340; Conservative 0; Mismatches 238; Indels 15; Gaps 1;

```
QY 1 GCCTCTACGCGGCGGCAACGCTTCTGACGCTTCCGCTGAGTGGCGAGT 60
DB 19618 GGGCGGTACGCGCGCGGTAGGCTTCTGAGGCTTCCGCTGACAGCCGGCGGAC 19677
QY 61 GGGCTGCGGTACCTGATGCTGCGGTGCTGCGCGCGGCAAAATGCGGTACC 120
DB 19678 GGGCGGCACTGACCTGCGGTGCGGTGCGCGCGGCAAGCGCGCTACCGAGGT 19737
QY 121 GAGGGGCGGCACTACCTGCGGCAAGCGCGGTGCGCGGCAATGACCGCGCGGCGATC 180
DB 19738 GCGACCGGAGAGCGCGCTGCGCGCTGCGCGCTGCGCGCGGCGCGCGCGCGCTC 19797
QY 181 GAGAGCTGCGGACCACTGCGGCAAGCGCGGCAAGCGGTGCGGTGCGGAC 240
DB 19798 ACCGCGGTGACACCGCGCTGCGCGCAAGCGCGCGGTGCGGTGCGGAC 19857
QY 241 CGGAGCGGTGCTGCACTGTCACCGCGCGCGCGCGCGCGCTTTCGACGAACTC 300
DB 19858 TGGTTCAGCTTCCCGCGCGCTTTCACCAAGCGCGCGCGCGCGCGCGCTTTCG 19917
QY 301 GGTGGGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345
DB 19918 CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19977
QY 346 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
DB 19978 CTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20037
QY 406 GTTCGAGCGGAGTGGCAAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 465
DB 20038 GTTCGAGCGGAGTGGCAAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 20097
QY 466 GTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 525
DB 20098 CGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 20157
QY 526 GCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 578
DB 20158 AAGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20210
```

RESULT 14  
US-09-320-878-19  
; Sequence 19, Application US/09320878A  
; Patent No. 6117659  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLIACH, Melanie C.  
; APPLICANT: BETLIACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 3062200210  
; CURRENT APPLICATION NUMBER: US/09/320,878A  
; EARLIER FILING DATE: 1999-05-27  
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
; EARLIER FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30  
; EARLIER APPLICATION NUMBER: 60/119,139

EARLIER FILING DATE: 1999-02-08  
; EARLIER APPLICATION NUMBER: 60/100,880  
; EARLIER FILING DATE: 1998-09-22  
; EARLIER APPLICATION NUMBER: 60/087,080  
; EARLIER FILING DATE: 1998-05-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19506  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-320-878-19

Query Match 28.7%; Score 172.2; DB 3; Length 38506;  
Best Local Similarity 57.3%; Pred. No. 6.8e-24;  
Matches 340; Conservative 0; Mismatches 238; Indels 15; Gaps 1;

```
QY 1 GCCTCTACGCGGCGGCAACGCTTCTGACGCTTCCGCTGAGTGGCGAGT 60
DB 17760 GGGCGGTACGCGCGGTAGGCTTCTGAGGCTTCCGCTGACAGCCGGCGGAC 17819
QY 61 GGGCTGCGGTACCTGATGCTGCGGTGCTGCGCGCGGCAAAATGCGGTACC 120
DB 17820 GGGCGGCACTGACCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 17879
QY 121 GAGGGGCGGCACTACCTGCGGCAAGCGCGGTGCGGTGCGGTGCGGTGCGGTGAC 180
DB 17880 GCGACCGGAGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 17939
QY 181 GAGAGCTGCGGACCACTGCGGCAAGCGCGGCAAGCGGTGCGGTGCGGTGCGGTGAC 240
DB 17940 ACCGCGGTGACACCGCGCTGCGCGCAAGCGCGCGGTGCGGTGCGGTGCGGTGAC 17999
QY 241 CGGAGCGGTGCTGCACTGTCACCGCGCGCGCGCGCGCGCGCTTTCGACGAACTC 300
DB 18000 TGGTTCAGCTTCCCGCGCGCTTTCACCAAGCGCGCGCGCGCGCGCGCTTTCG 18059
QY 301 GGTGGGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345
DB 18060 CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18119
QY 346 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
DB 18120 CTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18179
QY 406 GTTCGAGCGGAGTGGCAAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 465
DB 18180 GTTCGAGCGGAGTGGCAAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 18239
QY 466 GTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 525
DB 18240 CGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 18299
QY 526 GCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGAC 578
DB 18300 AAGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18352
```

RESULT 15  
US-09-141-908-1  
; Sequence 1, Application US/09141908  
; Patent No. 6503741  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLIACH, Melanie C.  
; APPLICANT: BETLIACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
; FILE REFERENCE: 3062200210  
; CURRENT APPLICATION NUMBER: US/09/141,908  
; EARLIER FILING DATE: 1998-08-28

```

? EARLIER APPLICATION NUMBER: CIP OF 09/073,538
? EARLIER FILING DATE: 1998-05-06
? EARLIER APPLICATION NUMBER: CIP OF 08/846,247
? EARLIER FILING DATE: 1997-04-30
? EARLIER APPLICATION NUMBER: PROV. 60/076,919
? EARLIER FILING DATE: 1998-03-05
? EARLIER APPLICATION NUMBER: PROV. 60/087,080
? EARLIER FILING DATE: 1998-05-28
? NUMBER OF SEQ ID NOS: 31
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1
? LENGTH: 38506
? TYPE: DNA
? ORGANISM: Streptomyces venezuelae
US-09-141-908-1

```

Query Match	28.7%;	Score 172.2;	DB 3;	Length 38506;
Best Local Similarity	57.3%;	Pred. No. 6.8e-24;		
Matches 340; Conservative	0;	Mismatches 238;	Indels 15;	Gaps 1;

QY 1 GCGCTCTACGAGGAGGGGCAACGCGCTTCTCTGAAAGCGCTTGGCCCGTGTGATGGGCGCAAT 60  
 Db 17760 GGCACGTAACGCGCGCGGTACGCGCTTCTTGAGCGCCTCGCGGTATGACACGGGCGCAAC 17819  
 QY 61 GAGCGTCGCGGTGACCTTGATCGACTCGGAGTCTGTGGAGCGGAGAGAACATGACCGGTACC 120  
 Db 17820 GGGCCCAACGTGACTCTGGTGGCTTGAAGCCCTCGGAGGGGAGAGCCGCTCACCGAGGAT 17879  
 QY 121 GAGGAGCGGCGACTTACTGCGCAAGCCAGAGGCTGCGCGCCATGAGCCCGAGCGGGCGATC 180  
 Db 17880 GCGACCGGAGAGCGGCTGCGCGCGCTCGGACTGCGCCCTCGCCCGCGAGCGGCGCTC 17939  
 QY 181 GAGGAGCTGCGGACCACTTGAGAGCGCGGGGACCCGTGGTGTGCGTGTGAGACTTGAC 240  
 Db 17940 ACCGCGCTGAGCAACCGCGCTGCGCAAGGCGGAGACACGCGCGTCCAGATCGCGACGTCGAC 17999  
 QY 241 CGGAGCGCGGTTGCTGCGAACTGTGTTCACCGCGCGCCCGCGCGGCGCCCTCTTCAGCAATC 300  
 Db 18000 TGGTGTGAGCTTTCGCCCCCGGCTTTACCAAGGCGCGGCGGAGCACTTCTTCGCGGATCTG 18059  
 QY 301 GGTGGGAGTCCGCGCGCGGCGCGGAGCA-----GACGCGTACAGAAATCGAAT 345  
 Db 18060 CCGGAGGCGCGCGCGCGCGCTGAGACGAGAGAGATGACGAGCGCGCGCGACGACCGTC 18119  
 QY 346 CTGCGCGCGCGGCTGCGCCTGCAATCGACCGGAGCGGAAATGTAAGAGCAATGTTCGCGGCTG 405  
 Db 18120 CTGAGCCCGCGAGCTCGGTGCGCTCAACCGGCGCGGACGAGAGCGCGTATGCGAGAGTTG 18179  
 QY 406 GTCCGAGCGGAGGTGAGCAAGCGGTCTGTGGGACCAAGGACCGCGGACCGGTGATCGAGCGTGAAC 465  
 Db 18180 GTCCGCGAGAGCACTTCGCGGTGTGCTTCAACCACTCCCTCCCGAGGCGGTGACACGGGG 18239  
 QY 466 GTTCGCTTTCGTTGACTTGGAATTGCACTTCATGACCGCGGTGACTTCGCGGAAACCGGCTTC 525  
 Db 18240 CGGCGCTTTCGTTGACTTCGATTTGCACTTCGCTGACCGCGGATCGAGCTTCGCAACCGCGCTC 18299  
 QY 526 GCGGCGGTGACCTGGGAGTCCGGGTGGCGACGACCACTGCTTTCACCAACCGCGAC 578  
 Db 18300 AAGGAACGCAACCGGCTTGCGCTCCCGGACCACTTCGTTCTTTCGACTTACCGGAC 18352

Search completed: April 7, 2006, 02:58:57  
Job time : 101.13 secs



## ORIGIN



Query Match	13.6%	Score 81.4	DB 10.4	Length 925
Best Local Similarity	13.2%	Pred. No. 2.2e-06		
Matches	52	Conservative 201	Mismatches 140	Indels 0
				Gaps 0
QY	145	CCGCGACGACGTGCGCTTACCGAGCTGCGCTTGCAGACCCCGAGACTCGAGCGAGTCTCC	204	
	533	SMSCWCKKCGCTTBSITTTTTSBSGSGYKGCSSBSGSGSBCSSCSGSSSSCCBCC	592	
Db	533	SMSCWCKKCGCTTBSITTTTTSBSGSGYKGCSSBSGSGSBCSSCSGSSSSCCBCC	592	
QY	205	TGCGCCGACGCGGTCTCTGTGTGTACACGAGCGCGCAGTACCCCCCGGCTGTGACGACGACTG	264	
	593	CCSCSSTCCSSBSSBSSSKCSTSTBSBCSCSSSKSYGVTSCSSSSSCSSSSSTSSSTSS	652	
Db	593	CCSCSSTCCSSBSSBSSSKCSTSTBSBCSCSSSKSYGVTSCSSSSSCSSSSSTSSSTSS	652	
QY	265	TCCCGACGCGCTTGAACACGCGCGCGCGACCGTGTCTTTGTGACCCGCGAGTGCACG	324	
	653	TSKSSSGSSSSSSSSSYTTSTKSTASGSGMSAAGGSGSTSTSSSSSSSTSSSSSV	712	
Db	653	TSKSSSGSSSSSSSSSYTTSTKSTASGSGMSAAGGSGSTSTSSSSSSSTSSSSSV	712	
QY	325	CCCGAGTCGCGCGCGCACTGCACGCGCGTGCACGCGACCGCCCTGTTCACGTGTCTC	384	
	713	SGSKSSTBSGSSBSSSGSSSSSTSSBBSCTSTSSSSSSSYBSPTCSCTCCCSYBSS	772	
Db	713	SGSKSSTBSGSSBSSSGSSSSSTSSBBSCTSTSSSSSSSYBSPTCSCTCCCSYBSS	772	
QY	385	TGCTTCGCGCTGACGAGCGCGAGTGTGTGCACGACCCGAGCTGACACCCCTCGGTTGG	444	
	773	TSSSSSTBSGSSBSSSVSTSSBBSSTPTCSGCCCTMCTCTCTYBMBCTTSSGSS	832	
Db	773	TSSSSSTBSGSSBSSSVSTSSBBSSTPTCSGCCCTMCTCTCTYBMBCTTSSGSS	832	
QY	445	TCCAGCGCGCTGCGCGCACCGGAGTTCACGTCGCCCTTGTGCTGTGTGACAGACGCG	504	
	833	SSSGKGYTKCGCGCGCGSSSTNGMBGTSAACSSSSSCSSSSSVSSSSSABSSSVSSG	892	
Db	833	SSSGKGYTKCGCGCGCGSSSTNGMBGTSAACSSSSSCSSSSSVSSSSSABSSSVSSG	892	
QY	505	CCGCGCTGACCGTTCGAGACGACGTCATCCGG	537	
	893	SSGVSNNSSSASKSSSGSVSSGSGSGSGSVG	925	
Db	893	SSGVSNNSSSASKSSSGSVSSGSGSGSGSVG	925	

RESULT 2	
BZ557931	
LOCUS	1438 bp DNA linear GSS 17-DEC-2002
DEFINITION	pacsl-60_6778.x2r5p11 pacsl-60 Pseudomonas aeruginosa genomic clone.
ACCESSION	BZ557931
VERSION	BZ557931.1 GI:27171396
KEYWORDS	GSS.
SOURCE	Pseudomonas aeruginosa
ORGANISM	Pseudomonas aeruginosa Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas. 1 (bases 1 to 1438) Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V. Whole-Genome-Sequene Variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol. (2002) In press Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@en.washington.edu Class: shotgun.
JOURNAL	
COMMENT	
TITLE	
REFERENCE	
AUTHORS	
FEATURES	
source	location/Qualifiers 1..1438 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="1-60" /db_xref="taxon:287" /clone_1=pacsl-60_6778" /clone_1ib=pacsl-60" /note="Clinical isolate 1-60 Whole genomic shotgun library."
ORIGIN	
Query Match	13.4%; Score 80.2; DB 9; Length 1438;
Best Local Similarity	47.8%; Pred.No. 3,7e-06;

[illegible]

```

RESULT 3
BU502934
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BU502934 1143 bp mRNA linear EST 12-SEP-2002
AGNCOURT 8929661 NIH_MGC_94 Mus musculus CDNA clone IMAGE:6489733
5', mRNA sequence.
BU502934
BU502934.1 GI:22808397
EST.
Mus musculus (house mouse)
Mus musculus
Bakaryot; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1143)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov

```

Plate: L1AM14038 row: 9 column: 14  
High quality sequence stop: 353.  
Location/Qualifiers  
1. 1143  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6489733"  
/issue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1db="NIH\_MGC\_94"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: Not1;  
Site 2: Sal1; Cloned unidirectionally; oligo-dT primed.  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 13.2%; Score 79.2; DB 5; Length 1143;  
Best Local Similarity 46.5%; Pred. No. 6e-06;  
Matches 224; Conservative 0; Mismatches 257; Indels 1; Gaps 1;

QY 86 CTACCCGCTTCCAAAGACAGAACTTCTGCTCCCGGTCCCTCGAGCGGGTCCCGACAC 145  
Db 375 CAGCCGNN 434  
QY 146 CGCGGAGAGATGAGCTTACAGCTCTGCTGAGACCCCTGAGACTCGGGGAGTCTCTCT 205  
Db 435 CAGCCGNN 494  
QY 206 GAGCCGAGCGAGTCTGCTGCTGAGACCGGAGGAGTACCCCGGCTGAGCGAGCTGAGT 265  
Db 495 GAGCCGNN 554  
QY 266 CCGCGAGCGGCTGAGAAAGCGCGGAGCGAGCTGCTGCTGAGACCGGAGTCTGAGT 325  
Db 555 GAGCGGCGGCGGAGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 614  
QY 326 CCGGAGTCTGAGCGGAGTCTGAGCGGAGTCTGAGCGGAGTCTGAGCGGAGTCTGAGT 384  
Db 615 GAGCGGCGGCGGAGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 674  
QY 385 TGCTCGGCTGAGCGGAGGAGGAGTCTGAGCGGAGTCTGAGCGGAGTCTGAGCGGAGT 444  
Db 675 CCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 734  
QY 445 TCAGAGCGCTGAGCGGAGCGGAGTCTGAGCGGAGTCTGAGCGGAGTCTGAGCGGAGT 504  
Db 735 CCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794  
QY 505 CCGCGGAGTCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 564  
Db 795 CCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 854  
QY 565 GG 566  
Db 855 GG 856

RESULT 4  
AG030591 888 bp DNA linear GSS 01-NOV-2001  
LOCUS AG030591  
DEFINITION Pan troglodytes DNA, clone: PTB-003A10.F, genomic survey sequence.  
ACCESSION AG030591  
VERSION AG030591.1 GI:16557464  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pan.  
REFERENCE  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

Totoki, Y., Watanabe, H., and Sakaki, Y.  
BAC end sequences of library PTB  
Unpublished  
2 (bases 1 to 888)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H., and Sakaki, Y.  
Direct Submission  
Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suenho-chou, Tsukuba, Ibaraki, Japan, URL: http://hsp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170  
Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the RFLP process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKS145  
R Site 1 : SacI  
R Site 2 : SacI

## COMMENT

Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the RFLP process and may have higher chance of  
clone tracking errors.

Sequencing: -21M13  
LIBRARY

Vector : pKS145  
R Site 1 : SacI  
R Site 2 : SacI

FEATURES  
source  
1. 888  
Location/Qualifiers  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-003A10.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_1db="PTB Chimpanzee Male BAC Library"

## ORIGIN

Query Match 12.6%; Score 75.6; DB 10; Length 888;  
Best Local Similarity 47.4%; Pred. No. 3.2e-05;  
Matches 260; Conservative 0; Mismatches 285; Indels 3; Gaps 2;

QY 1 CCGCGTCTGCGGAGGAGTACACCGGAGGAGTGAAGTCACTGCGGAGTCACTGCGGAGTGA 60  
Db 319 CCGCGGAGTCTGCGGAGGAGTGAAGTCACTGCGGAGTCACTGCGGAGTGAAGTCACTG 378  
QY 61 AGGAGAGCGGCGGAGTGAAGTCACTGCGGAGTGAAGTCACTGCGGAGTGAAGTCACTG 120  
Db 379 CAGCCGNN 438  
QY 121 TCCCGTGGAGCGGAGTCTGCGGAGTGAAGTCACTGCGGAGTGAAGTCACTGCGGAGT 180  
Db 439 CCGCGGAGTCTGCGGAGTGAAGTCACTGCGGAGTGAAGTCACTGCGGAGTGAAGTCA 498  
QY 181 CCGTGAAGTCTGCGGAGTGAAGTCACTGCGGAGTGAAGTCACTGCGGAGTGAAGTCA 240  
Db 499 CCGCGGAGTCTGCGGAGTGAAGTCACTGCGGAGTGAAGTCACTGCGGAGTGAAGTCA 558  
QY 241 TACCCCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300  
Db 559 GCGCGGAGTCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 618  
QY 301 TGTGTGACCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 360  
Db 619 CCGCGGAGTCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 677  
QY 361 CCGCGGAGTCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 420  
Db 678 CCGCGGAGTCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 737  
QY 421 CCAAGCTGAGAC--CTGCGTGTGTCAGAGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 478  
Db 738 GCGCGGAGTCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 797  
QY 479 CCGTGTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 538  
Db 798 GCGCGGAGTCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 857  
QY 539 CCAAGCTGA 546

Db	858	CANNCCCA	865
RESULT 5			
CNS0091P/c			
LOCUS			
DEFINITION	CNS0091P	925 bp	DNA linear GSS 03-JUN-1999
ACCESSION	Drosophila melanogaster genome survey sequence TET3 end of BAC #		
VERSION	BACR1916 of RPCI-98 library from Drosophila melanogaster (fruit		
KEYWORDS	fly), genomic survey sequence.		
SOURCE	AL053013		
ORGANISM	AT053013.1	GI:4934461	
REFERENCE	GSS.		
AUTHORS	Drosophila melanogaster (fruit fly)		
JOURNAL	Drosophila melanogaster		
COMMENT	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
	1 (bases 1 to 925)		
	Genoscope.		
	Direct Submission		
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : segrefigenoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
	Determination of this BAC-end and sequence was carried out as part of a		
	collaboration with the Berkeley Drosophila Genome Project (BDGP).		
	The BDGP is constructing a physical map of the Drosophila		
	melanogaster genome using these BACs. For further information		
	please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila		
	melanogaster BAC library was prepared by Kazuhiro Osoegawa and		
	Aaron Mamonos in Pieter de Jong's laboratory in the Department of		
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,		
	NY. The library is named RPCI-98 and was constructed by partial		
	Scori digestion of Drosophila DNA provided by the BDGP from the		
	isogenic strain Y2, cn bw sp, the same strain used for the BDGP's		
	P1 and EST libraries. A more detailed description of the library		
	and how to order individual BAC clones, the entire library, or		
	filters for hybridization from the BACpac Resource Center can be		
	found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .		
FEATURES	location/Qualifiers		
source	1..925		
	/organism="Drosophila melanogaster"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:7227"		
	/clone="BACR1916"		
	/clone_1lb="RPCI-98"		
	/note="end : TET3"		
ORIGIN			
Query Match	12.5%	Score 75.2;	DB 10; Length 925;
Best Local Similarity	13.3%;	Pred. No. 3.8e-05;	
Matches	52;	Conservative 194;	Mismatches 144; Indels 0; Gaps 0;
164	CGAGTTCGCTGGACACCCCGTAGACTCGGGCGGCTTCCTCGCGCGACGGCTCTGT	223	
921	CSCGCSBSCBSGSSSSMTSSNSBSCSSBSSSSTSSSSSSSSSSSSSSGCT	862	
224	GATGACCGGAGCGGAGTACCCCGAGCTGACGAGACGTCGTCGCGCTGAACA	283	
861	SSATCTKCAASSCGCGCGCAGBACMCSSSSSCGASAGVAVASGAGKRGSSGAS	802	
284	GGCGGGGCGACCGTGTGTGTGACACCGAGATGTCGCGCGCCCGAGATCGCGCACT	343	
801	ASHSSSSSCBSSSSSCSASCMASSSSSASRSRSGGAGGASSSRSSSSSASAG	742	
344	CGACCCCGTCGACGCGACCGCCTGTGCACTGTGCTCTGCTGCGCTGCGCGAGG	403	
741	SVVSSSSSSSSSSSVSCSVASMSGSSBSSSSASSSSSSSASGASCCCTTWS	682	
404	CGGTCTGTGACGACCCCGAGCTGACACCTTCGCTTGTGTCACAGCGCTCGCGAC	463	
681	CGCTGASMSAARSSSSSSSCSSSSMSASASSSSSSSSSSSSSGAGACGMSMS	622	

Oy		461	CGGATGATGACGTCCCTCCTGGTGCCTGTGTGAACAAGGAGACGCCGCCGCGTGACCTGTTCGAGA	523
Dd		621	GCGSGGSVASSGMSVVSSSGGRSBSGGGGGCTGAGSSBSGSGSVCSGSCG	562
Oy		524	CGACGTGCATCCGGCCAGCCCATGTTCG	553
Dd		561	CRCCSSAAAAAASCVAAACGMCMGKKSMSG	532
<hr/>				
RESULT 6 CNS006XK/c				
LOCUS				
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #				
BACR1AN09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION AL066051				
VERSION AL066051.1 GI:4945019				
KEYWORDS GSS.				
SOURCE Drosophila melanogaster (fruit fly)				
ORGANISM Drosophila melanogaster				
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE 1 (bases 1 to 935) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : Submitted BP 191 91006 Evry cedex - FRNACE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the logenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.				
FEATURES				
source location/Qualifiers				
1..935				
/organism="Drosophila melanogaster"				
/mol_type="genomic DNA"				
/_db_xref="taxon:7227"				
/clone="_BACR1AN09"				
/_clone_1lb="RPCI-98"				
/_note="end : T7"				
<hr/>				
ORIGIN				
Query Match 12.5%; Score 74.8; DB 10; Length 935;				
Best Local Similarity 30.2%; Pred. No. 4.6e-05;				
Matches 125; Conservative 114; Mismatches 172; Indels 3; Gaps 1				
Oy		112	GGCTCCCGCTCCCCTCGGCGGCGGTCCCCGACACCGGAGCAGAGTGGGCGTTAACAGCTCG	171
Dd		906	GSCTCCSGSCCCSGCGCGCCSSCGSSSCCGGSCCGGSCCGGSSCCSGCGCGCGS	847
Oy		172	CCTGTGACACCCGTCGACTCGGCGGATCTCTCCCTGTGACCAGACGAGTCTCTGTGTGTGACCG	231
Dd		846	SAGCGCCCGSGSGCGCCCGSCGSGSCGSSCCGCGCGCGSCGCGCCGCGCGCGCGGCG	787
Oy		232	GAGCGGCAGTACCCCGGCGCTTGACGACGAGTGTTCGCGACGCGCTTGAAACAGCGCGG	291
Dd		786	SSGCGCGCGCGCGSCCGCGSKCGCGCGSGGSCGSSSGSGSCGSGCGSCGSCGCS	727
Oy		232	CGACCGTGTGTGTGTGACACCGGCGAGTGTGCGCGCGCCGATGCGCGCCCGCACATCGACGCG	351
Dd		726	CGCGSGGCGCGCGCG--CGCGCGCGCGCGSGCGSGCGCGCGCGCGCGCGCGCGCG	670

[illegible]

RESULT 7	
B0922913	
LOCUS	B0922913
DEFINITION	1169 bp mRNA linear EST 20-AUG-2002
AGENCOURT	NCI CGAP Co24 Muv musculus cDNA clone
IMAGE:	6396111 5', mRNA sequence.
ACCENSOR	D000000000

ORGANISM	SOURCE	KEYWORDS	VERSION	ACCESSION
Mus musculus	Mus musculus (house mouse)	EST.	GI:22337944	B0242913
Mus musculus				B0242913

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1169)	NIH-MGC	<a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a>		
	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strusberg, Ph.D.			

Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNI)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNI at:  
<http://image.llnl.gov>  
Plate: LMNI3891 row: n column: 16  
High quality sequence stop: 147.

FEATURES	Location/Qualifiers
SOURCE	1. .1169

```

/organism="Mus musculus"
/mol type="mRNA"
/strain="FVB/N"
/db xref="taxon:10090"
/clone="IMAGE:639611"
/lab host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_C024"
(note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by life
technologies. Note: this is a NCI CGAP library.")

```

Query Match	12.2%	Score 73;	DB 5;	Length 1169;
Best Local Similarity	46.7%	Pred. No. 0.00011;		
Matches 279;	Conservative	0;	Mismatches 313;	Indels 6;
			Gaps	2;

Qy	2	CGCCCTGCGCCGAGGCGTACACCCCGGGGGGGTGAAGTCCATCGCGTACCGCATGGGGTGA	61
Db	202	CGCGNCGGGCGGGGGGGGGGGCGCGGGGGGGGGCGGGGGCCCGGGGGGGGCGCG	261
Qy	62	GGGACGCGCCGGTGCACGTCGCGGTCACCGGTTCAACGACAGAACTTCTGGCTCCGGT	121
Db	262	CGGCGGGCGCGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGGGGGGGGGCGGGCG	321
Qy	122	CCCCCTGGGCGGGGTCCCGGACACCGGCGGACGATGCGCGTTACCAAGTTCGCGTGGCACCC	181
Db	322	CNCGCGCGCGCCCGCGNCGCCCGCGGGCGCGCGCGGCGCGCCCGCGCGGGGGGGCGCG	381

Oy 182 CGTGAACCTGGAGAGGCTCTCCCTGGGCGGAGACGAGATCTATGTAGTGAACCGAGGCGAGT 241  
 Db 382 GGGCGGAGGGGCGCGGGGCGGCGGGGGGGGGGCGCCGCGCGGGCGCGCGCGGGCGG 441  
 Oy 242 ACCCCGCGGACCTGAGACGAGACGTGGTCCGCGACGGCTGTAGAACAGCGCGGGCGACCGTCT 301  
 Db 442 GGGGCGGGGCGGGGGGCGCGGGGGCGGGGCGCGCCCGCGGGCGCGGGGGGGGGGCGCGGG 501  
 Oy 302 GTTGTGC--ACCGGCGAGTGGCGCGCGCCCGGATCGAGCGCGCGACTCGACGCGGTTCAGCGCG 359  
 Db 502 GGGCGGGGAGCGCGGCGCGCGCGCGCGCGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGGG 561  
 Oy 360 ACCGCGCTGTCACTGTAAGTCTCTGTCTGTGGCGCGCGCGAGGAGCGGTGTGTGTCAACGAC 419  
 Db 562 GGGGCGCCCGCGGGGCGCGCGGGGGGGGGGCGCGCGCGCGCGCGGGGCGGGGGCGG 621  
 Oy 420 CCCAGCCTGAGCAACCTCGGTTGTGTCAAGCGCTCGCGCGACGCGGAGTCAGTCTCC 479  
 Db 622 GGGCCCGGCGCGGGGCGCGGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 677  
 Oy 480 CTGTGGCTGGTGAACCAAGGAGCGCGCGCGCGCTGTACCGTCCGAAGACAGCTCGATCCGAGC 539  
 Db 678 GCCGANNCGGAGGCGCCCG 737  
 Oy 540 CAGGCGCATGGTCGGGTGGGCTCGGCGGGGATGTATGGAGGCTGGAAGTCCCGCGCGCGGTGG 597  
 Db 738 CGGGCGGGGCGCGGGGCGCGCGGGGAGGAGGCGGGGCGCGCCCGCGCGGGGAGGCGGGG 795

RESULT 8	AG060114/c	852 bp	DNA	linear	GSS 03-NOV-2001
LOCUS	AG060114				
DEFINITION	Pan troglodytes DNA, clone: PTB-047105.F, genomic survey sequence.				
ACCESSION	AG060114				
VERSION	AG060114.1	GI:16611344			
KEYWORDS	GSS.				
SOURCE	Pan troglodytes (chimpanzee)				
ORGANISM	Pan troglodytes				

REFERENCE AUTHORS	TITLE JOURNAL	REFERENCE AUTHORS	TITLE JOURNAL
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of library P1B	2 (bases 1 to 852) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	Unpublished
Direct Submission Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical			

**COMMENT**  
Clones are derived from the chimpanzee BAC library PB3 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

```

Sequencing: -21M13
LIBRARY
Vector      : PKS145
R.Site 1    : SacI
R.Site 2    : SacI
FEATURES
Source
1. 852
   /organism="Pan troglodytes"
   /mol_type="genomic DNA"
   /db_xref="taxon:9598"
   /clone="PTB-047105.F"
   /sex="male"
   /cell_type="lymphoblast"
   /clone_lib="PTB Chimpanzee Male BAC Library"

```

## ORIGIN

Query Match 12.1%; Score 72.8; DB 10; Length 852;  
Best Local Similarity 46.0%; Pred. No. 0.00012;  
Matches 267; Conservative 0; Mismatches 308; Indels 6; Gaps 1;

QY 1 CCGGCTCCGCGGAGGCTGACCCGCGGAGGAGGCTGACCTGCGGAGGAGGAGG 60  
DB 831 CCGGCTCCGCGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGG 772  
QY 61 AGGAGCCGCGGAGGCTGACCTGCGGAGGCTGACCTGCGGAGGAGGAGGAGG 120  
DB 771 CCGGCTCCGCGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGG 718  
QY 121 TCCGCTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGG 180  
DB 717 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGG 658  
QY 181 CCGTGAACCTGCGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGG 240  
DB 657 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGG 598  
QY 241 TACCCGCGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGG 300  
DB 597 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 538  
QY 301 TGTGTGACCGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGG 360  
DB 537 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 478  
QY 361 CCGGCTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 420  
DB 477 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 418  
QY 421 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 480  
DB 417 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 358  
QY 481 TGTGTGACCGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 540  
DB 357 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 298  
QY 541 AGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 581  
DB 297 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 257

RESULT 9 CA981964 985 bp mRNA linear EST 27-FEB-2003  
CA981964  
LOCUS AGNCOURT 11280952 Wellcome CRC PSK egg Xenopus laevis cDNA clone  
DEFINITION IMAGE:6870527 5', mRNA sequence.

ACCESSION CA981964  
VERSION CA981964.1 GI:27514618  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM

REFERENCE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLES Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rtm1.nih.gov  
Tissue Procurement: N. Garrett, P. Lemaire, A.M. Zorn, and J.B.  
Gordon (Wellcome/CRC Institute)  
cDNA Library Preparation: N. Garrett, P. Lemaire, A.M. Zorn, and  
J.B. Gordon (Wellcome/CRC Institute)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: XGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ULNL at:  
http://image.jnl.gov  
Plate: LLAM4501 row: a column: 22  
High quality sequence stop: 241.  
Location/Qualifiers

## FEATURES

source

1..985

/organism="Xenopus laevis"

/mol\_type="mRNA"

/db\_xref="taxon:8355"

/clone="IMAGE:6870527"

/issue\_type="egg"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="Wellcome CRC PSK egg"

/notes="Vector: Bluescript SK-; Site 1: NotI; Site 2:  
EcoRI; cDNAs were oligo-dT primed and directionally  
cloned. Library was constructed by N. Garrett, P. Lemaire,  
A.M. Zorn, and J.B. Gordon (Wellcome/CRC Institute)."  
Note: This is a Xenopus Gene Collection (XGC) library."

## ORIGIN

Query Match 12.1%; Score 72.8; DB 6; Length 985;  
Best Local Similarity 46.9%; Pred. No. 0.00012;  
Matches 250; Conservative 0; Mismatches 282; Indels 1; Gaps 1;

QY 35 GGTGACCTGCGGAGGCTGACCCGCGGAGGAGGCTGACCTGCGGAGGAGGAGGAGG 94  
DB 343 GGTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 402  
QY 95 CCAAGCAGAGAACTTCTGAGCTCCGCGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 154  
DB 403 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 462  
QY 155 GTGAGCTTACGAGCTGCGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGG 214  
DB 463 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 522  
QY 215 GGTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 274  
DB 523 GGTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGG 582  
QY 275 CTTGAGAAAGAGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGG 334  
DB 583 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 642  
QY 335 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 394  
DB 643 GGTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 701  
QY 395 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 454  
DB 702 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 761  
QY 455 GCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 514  
DB 762 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 821  
QY 515 GGTGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 567  
DB 822 CCGGAGGAGGAGGCTGACCCGCGGAGGAGGAGGCTGACCTGCGGAGGAGGAGGAGGAGGAGGAGG 874

## RESULT 10

AG159111

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AG159111 1003 bp DNA linear GSS 09-JAN-2002  
Pan troglodytes DNA, clone: RP43-024116.T7, genomic survey  
sequence.  
AG159111  
AG159111.1 GI:16688789  
GSS.  
Pan troglodytes (chimpanzee)  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





[illegible]

RESULT	12
AG076818	
LOCUS	AG076818 1152 bp DNA linear GSS 03-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-071C05.R, genomic survey sequence.
ACCESSION	AG076818
VERSION	AG076818.1 GI:16628620
KEYWORDS	GSS.
SOURCE	Pan troglodytes (chimpanzee)
ORGANISM	Pan troglodytes

ORGANISM Pan troglodytes  
Euleryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eucarchontia; Eumammalia; Primates; Catarrhini;  
Hominoidea; Pan.  
Homidae; Pan.

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
1	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of library PTB	Unpublished	2 (bases 1 to 1152)	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.

**JOURNAL** Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suheiyo-Chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimpbesegsc.riken.go.jp](mailto:chimpbesegsc.riken.go.jp), URL: <http://ngp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)

**COMMENT** Clones are derived from the chimpanzee BAC library PIB. This BAC end was generated during the Red process and may have higher chance of clone tracking errors.

```

Sequencing: M13Rev
LIBRARY
Vector      : pRS145
R.Site 1    : SacI
R.Site 2    : SacI
Location/Qualifiers
1. .1152
FEATURES
source

```

```

/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

```

Query Match	12.0%	Score 71.8;	DB 10;	Length 1152;
Best Local Similarity	48.4%;	Pred. No. 0.00018;		
Matches 237; Conservative	0;	Mismatches 250;	Indels 3;	Gaps 2;

QY	113	GCATCCGAGTCCCTGAGGCGAGGATCCCGACACACCGAGGAAAGAAATGAGGTTACAGACGTCCG	172
Db	630	GCGCGCGAGGCGACATGCCCCCGCGCCCCCGCGCGCCCCCGGNNNGAGGCGCGCGCCG	689
QY	173	CTGACACCCCGCTGACTCGAGGAGGTCTTCCCTGACCGAACGGGTCTTGATGATGAAACCG	232
Db	690	CGGCGCGGCGNCCCGCCCGCGCCCGNCGCCCCCGCGCGCGCGCGAGCCGCGCCGCG	749
QY	233	AGCGGACGTAACCCCGGAGCTGAGACGGAAGTGGATCCGCGACAGGACTCTGAGACAGACGCGAGCG	292
Db	750	CGGCGACGCGAGCGN	809
QY	293	GACCGTGTGTTGTGACCGCGGAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	352
Db	810	GCCGCGCGAGGAGCGAGCG	869
QY	353	CGACGAGACCGGCGCGCTGTCACTGTGTGTCTCTGTGCTGAGCGCTGCGCGAGGAGCGTGTGTGT	412
Db	870	CG	928
QY	413	CGACGACCCCGACGCTGAGACACTCGCGTTGATCAGAGCGCTCGGCGCGAGCGGAGATCGA	472
Db	929	GCGCGGAGCGCGCGCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCG	988
QY	473	CGTCCCGCTGTGAGCTGTGTGACCAAGAGACGCGCGCGCGCTGACCGCTCGAGAGACGATCGA	532
Db	989	CGAGCGCCCG	1048
QY	533	TCCGCGCCAGGCGCATGTGCGGTGGG--TCGCGCGGATGTGAGCGGTGAGATGCCCGCGCC	590
Db	1049	GCGGCGCGAGCGCGCGCGGAGAGGAGCGCGCGCGCGCGGAGGAGGAGCGCGCGCGCGCGCG	1108
QY	591	CGGTGGGATG 600	
Db	1109	CGCGCGGAGCG 1118	

RESULT 13	
LOCUS	AG072414
DEFINITION	996 bp DNA linear GSS 03-NOV-2001
ACCESSION	Pan troglodytes DNA, clone: PTB-063011.R, genomic survey sequence.
VERSION	AG072414
KEYWORDS	AG072414.1 GI:1624216
SOURCE	GSS.
ORGANISM	Pan troglodytes (chimpanzee)
	Pan troglodytes
	Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Pan.

REFERENCE	AUTHORS	TITLE
1	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of library PB8

JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical  
 Direct Submission  
 Totsuki, Y., Watanabe, H. and Sakaki, Y.  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 2 (bases 1 to 996)  
 Unpublished

COMMENT

and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Shuhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimbases@gsc.riken.go.jp](mailto:chimbases@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PB8. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

Sequencing: M13Rev  
LIBRARY



Vector : PKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .996  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-063011.R"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC library"

## ORIGIN

Query Match 11.9%; Score 71.6; DB 10; Length 996;  
Best Local Similarity 46.9%; Pred. No. 0.0002;  
Matches 271; Conservative 0; Mismatches 301; Indels 6; Gaps 2;

25 GGGGGGTGAGTCTGACCTGATCCGCACTGGGTGAGGAGCCCGCTGACCTGCGG 84  
|||  
405 GGGGGGGCGCGCGGANNCGCGCGGGGAGGCGGGCCCCCGCGCGCGCGCGCG 464  
|||  
85 TCTACCGCTTCCACGACAGAACTTCTGCTCCGCTCCGCTGCGGCGGCTCCGACA 144  
|||  
465 GCG 524  
|||  
145 CCGGCGACGAGTGGCGCTTACCACTGCGCTGCGACCCCGCTGACCTTGGCGGCTCTCC 204  
|||  
525 CCGGCG 584  
|||  
205 TGGCGGAGCGGCTCTGCTGCTGCTGAC--CGAGGCGGAGTACCCCGCGCTGAGAGT 262  
|||  
585 CG 644  
|||  
263 GGTCCCGACGCGCTTGAAACGCGCGGCGGACCGCTGCTTGTGTGACCGCGAGTGC 322  
|||  
645 CCGGCG 704  
|||  
323 CGCGCGGATCGGCG 382  
|||  
705 CG 764  
|||  
383 TCTGCTCGCGCTGCGCGAGGCGGCTGCTGACGACCCCGAGCTGAGACCCCTCGGTT 442  
|||  
765 GCG 820  
|||  
443 GGTCCAGCGCTTGCAGCGCGCGGATTCGACGTCCCGCTGCTGCTGCTGACGAGAGCG 502  
|||  
821 GCG 880  
|||  
503 CGCGCGCGCTGACCTGCTGAGAGACGCTCGATCCGCGCCAGGCGATGCTGCTGCTCG 562  
|||  
881 CG 940  
|||  
563 CGCGGCTGCTGCGCGGCTGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
|||  
941 GCG 978  
|||

RESULT 14  
AG039543/c 1101 bp DNA 1linear GSS 01-NOV-2001  
LOCUS AG039543  
DEFINITION Pan troglodytes DNA, clone: PTB-016K22.F, genomic survey sequence.  
ACCESSION AG039543  
VERSION AG039543.1 GI:16568268  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homidae; Pan.  
REFERENCE  
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of library PTB  
Unpublished  
2 (bases 1 to 1101)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suenhiro-chou, Tsukuba, Ibaraki, Japan, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9111, Fax: 81-45-503-9170  
Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the R&D process and may have higher chance of  
cloning errors.  
PRIMERS  
Sequencing: -21M13

## COMMENT

LIBRARY  
Sequencing: -21M13

Vector : PKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .1101  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-016K22.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC library"

FEATURES  
source

## ORIGIN

Query Match 11.9%; Score 71.6; DB 10; Length 1101;  
Best Local Similarity 46.4%; Pred. No. 0.0002;  
Matches 224; Conservative 0; Mismatches 257; Indels 2; Gaps 1;

118 CGGTCCCGCTGCG 177  
|||  
892 CG 833  
|||  
178 ACCCGCTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 237  
|||  
832 GCG 773  
|||  
238 CAGTACCCCGCGCTGAGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297  
|||  
772 CACCG 713  
|||  
298 TGTGTGTGTGACCG 357  
|||  
712 GCG 653  
|||  
358 GACCGCGCGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 417  
|||  
652 GCG 593  
|||  
418 ACCCGAGCTGAGACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477  
|||  
592 GCG 535  
|||  
478 CCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537  
|||  
534 CCGGCG 475  
|||  
538 CCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597  
|||  
474 GCG 415  
|||  
598 GTG 600  
414 GCG 412

RESULT 15  
AG332984/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
ORIGIN

AG332984 1440 bp DNA linear GSS 18-DEC-2004  
Mus musculus molossinus DNA, clone:MSMg01-123F20.T7, genomic survey  
sequence.  
AG332984 GI:47906294  
GSS.  
Mus musculus molossinus (Japanese wild mouse)  
Mus musculus molossinus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1  
Abd, K., Noguchi, H., Tagawa, K., Yuzurika, M., Toyoda, A., Kojima, T.,  
Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriaki, K. and  
Shiroishi, T.  
Contribution of Asian mouse subspecies Mus musculus molossinus to  
genomic constitution of strain C57BL/6J, as defined by BAC-end  
sequence-SNP analysis  
Genome Res. 14 (12), 2439-2447 (2004)  
15574823  
2 (bases 1 to 1440)  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
Direct Submission  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail:hattori@sc.riken.jp, URL:http://hsp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
Taishu Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp  
PRIMERS  
Sequencing : T7  
LIBRARY  
Vector : pBACe3.6  
R.site 1 : EcoRI  
R.site 2 : EcoRI.  
Location/Qualifiers  
1..1440  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-123F20.T7"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"

Query Match 11.9% Score 71.6; DB 10; Length 1440;  
Best Local Similarity 47.0%; Pred. No. 0.0002;  
Matches 254; Conservative 0; Mismatches 285; Indels 2; Gaps 1;

QY 1 CCGCGCTGCGCGAGCGTACACCCGCGGCGTGGAGTCACTGCGGCTACCGGAGTGGTG 60  
DB 1062 CCCCCCTCG 1003  
QY 61 AGGAGCGCCCGGTGCACTGCGCGGTCTACCGGTTCAACGACAGACTTTTGGCTCCGG 120  
DB 1002 CGG--CGGCG 945  
QY 121 TCCCCCTGGGCGCGGTCCCGGACACGGGAGAGTGGGCTTACAGGCTGGCTGGACG 180  
DB 944 CCG 885  
QY 181 CCGTGAAGCTCGGCGGTCTCTCCCTGCGCGGACGCGGTCTTGGTGTGACCGGAGCGGCG 240

DB 884 CCG 825  
QY 241 TACCCCGCGCTGAGACGAGCGTGTCTCGACGAGCGCTGAAACAGCGCGGCGGACCGCTG 300  
DB 824 GCG 765  
QY 301 TGTGTGACACCGCGCAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360  
DB 764 CCG 705  
QY 361 CCG 420  
DB 704 CCG 645  
QY 421 CCAAGCTGACACCGCTGCGCTTGTGATCCAGCGCGCTCGCGCGCGCGCGCGCGCGCGCG 480  
DB 644 CCG 585  
QY 481 TGTGTGCTGTGACCGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
DB 584 CCG 525  
QY 541 A 541  
DB 524 A 524

Search completed: April 7, 2006, 14:04:00  
Job time : 2314.57 secs





```
FT /gene= "mega"
FT /function= "KS1"
FT misc_feature
FT /tag= p
FT /gene= "mega"
FT /function= "AT1"
FT misc_feature
FT /tag= q
FT /gene= "mega"
FT /function= "KR1"
FT misc_feature
FT /tag= r
FT /gene= "mega"
FT /function= "ACp1"
FT misc_feature
FT /tag= b
FT /gene= "mega"
FT /function= "KS2"
FT misc_feature
FT /tag= t
FT /gene= "mega"
FT /function= "AT2"
FT misc_feature
FT /tag= u
FT /gene= "mega"
FT /function= "KR2"
FT misc_feature
FT /tag= v
FT /gene= "mega"
FT /function= "ACP2"
FT CDS
FT /tag= w
FT /gene= "megaII"
FT /product= "megalomycin 6-deoxyerythronolide B synthase 2"
FT /note= "Polyketide synthase, encodes AAB82213"
FT misc_feature
FT /tag= x
FT /gene= "megaII"
FT /function= "KS3"
FT misc_feature
FT /tag= ai
FT /gene= "megaII"
FT /function= "KS5"
FT misc_feature
FT /tag= y
FT /gene= "megaII"
FT /function= "AT3"
FT misc_feature
FT /tag= aj
FT /gene= "megaII"
FT /function= "AT5"
FT misc_feature
FT /tag= z
FT /gene= "megaII"
FT /function= "KR3 (inactive)"
FT misc_feature
FT /tag= ak
FT /gene= "megaII"
FT /function= "KR5"
FT misc_feature
FT /tag= aa
FT /gene= "megaII"
FT /function= "ACP3"
FT misc_feature
FT /tag= al
FT /gene= "megaII"
FT /function= "ACP5"
FT misc_feature
FT /tag= ab
FT /gene= "megaII"
FT /function= "KS4"
FT misc_feature
FT /tag= am
```

```
FT /gene= "megaII"
FT /function= "KS6"
FT misc_feature
FT /tag= ac
FT /gene= "megaII"
FT /function= "AT4"
FT misc_feature
FT /tag= an
FT /gene= "megaII"
FT /function= "AT6"
FT misc_feature
FT /tag= ad
FT /gene= "megaII"
FT /function= "DH4"
FT misc_feature
FT /tag= ao
FT /gene= "megaII"
FT /function= "KR6"
FT misc_feature
FT /tag= ae
FT /gene= "megaII"
FT /function= "BR4"
FT misc_feature
FT /tag= ap
FT /gene= "megaII"
FT /function= "ACP6"
FT misc_feature
FT /tag= af
FT /gene= "megaII"
FT /function= "KR4"
FT misc_feature
FT /tag= ag
FT /gene= "megaII"
FT /function= "TB"
FT misc_feature
FT /tag= ag
FT /gene= "megaII"
FT /function= "ACP4"
FT CDS
FT /tag= ah
FT /gene= "megaII"
FT /product= "megalomycin 6-deoxyerythronolide B synthase 3"

Query Match 100.0%; Score 601; DB 4; Length 47981;
Best Local Similarity 100.0%; Pred. No. 1,6e-91;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTCCTACGCGCGGCAACGCTTCCTGACGCGCTTGCGCGTGTGTGGCGCAGT 60
DB 37506 GCCTCCTACGCGCGGCAACGCTTCCTGACGCGCTTGCGCGTGTGTGGCGCAGT 37565
QY 61 GGGCTGCGCGTCACTCGATCGCTGTGGGTCTGTGGGCGGCAAGACATGCGCGTACC 120
DB 37566 GGGCTGCGCGTCACTCGATCGCTGTGGGTCTGTGGGCGGCAAGACATGCGCGTACC 37625
QY 121 GAGGCGGCGACTACCTTGCGCAGCGAGGCTTGCGCGCCATGACCGCGAGCGGCGATC 180
DB 37626 GAGGCGGCGACTACCTTGCGCAGCGAGGCTTGCGCGCCATGACCGCGAGCGGCGATC 37685
QY 181 GAGGAGCTGCGGACCAACCTTGAGCGCGGCGGACCGGTGGTGTGGTGTGACCTGAGC 240
DB 37686 GAGGAGCTGCGGACCAACCTTGAGCGCGGCGGACCGGTGGTGTGGTGTGACCTGAGC 37745
QY 241 CGGAGCGGTTCTGTGAATCTGTTCACCGCGCGCGCGCGCCCTCTTTCAGACGATC 300
DB 37746 CGGAGCGGTTCTGTGAATCTGTTCACCGCGCGCGCGCGCCCTCTTTCAGACGATC 37805
QY 301 GGTGGGTTCCGCGCGGCGCGGCGGAGACCGGTCAAGAAATCGGAATCTGCGCGGCGCTG 360
DB 37806 GGTGGGTTCCGCGCGGCGCGGCGGAGACCGGTCAAGAAATCGGAATCTGCGCGGCGCTG 37865
QY 361 GCGTCGATCCCGAGGCGCAACGTCAAGAGCATGTGCGCGGCTGAGTCCGAGCCGAGTGT 420
```

```
D 37866 GCGTCATGCCGAGGCGGAACGTACAGACAGATGCGCCGGCTGTCCGAGCGAGGTG 37925
Q 421 GCACGGGTGCTGGGCGCCACGGGACACCGGTGATCGAGGGTGAAGCTGCGCTTCCGCGAC 480
D 37926 GCAGCGGTGCTGGGCGCCACGGGACACCGGTGATCGAGGGTGAAGCTGCGCTTCCGCGAC 37985
Q 481 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTGGCGGAAACCGGCTCGCGGCGGTGACCGGG 540
D 37986 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTGGCGGAAACCGGCTCGCGGCGGTGACCGGG 38045
Q 541 GTCCGGGTGGCGACGACCATGCTTTTGACCAACCGGACAGTGAACCGGCTTCGCGCGGAC 600
D 38046 GTCCGGGTGGCGACGACCATGCTTTTGACCAACCGGACAGTGAACCGGCTTCGCGCGGAC 38105
Q 601 T 601
D 38106 T 38106

RESULT 3
AAQ46806
ID AAQ46806 standard; DNA; 29879 BP.
XX
AC AAQ46806;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 22-DEC-1993 (first entry)
XX
DE eryA region of S. erythraea chromosome.
XX
KW Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;
KM erythromycin; condensation; elongation; acyl chain growth;
KN gene replacement; ss.
XX
OS Saccharopolyspora erythraea.
XX
FH Key Location/Qualifiers
FT CDS 3..10139
FT FT /*tag= a
FT FT /note= "ORF 1"
FT CDS 10218..20921
FT FT /*tag= b
FT FT /note= "ORF 2"
FT CDS 20922..29879
FT FT /*tag= c
FT FT /note= "ORF 3"
XX
PN MO9313663-A1.
XX
PD 22-JUL-1993.
XX
PF 17-JAN-1992; 92WO-US000427.
XX
PR 17-JAN-1992; 92WO-US000427.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Katz L, Donadio S, Mcalpine JB;
XX
DR MPI: 1993-242804/30.
DR P-PSDB; AAR44430, AAR44431, AAR44432.
XX
PT Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -
PT by introducing altered biosynthetic gene-contg. DNA into microorganisms.
XX
PS Claim 27; Fig 2; 133pp; English.
XX
XX This sequence represents a fragment of the Saccharopolyspora erythraea
CC genome, designated eryA. The polypeptides encoded by this region are
CC involved in the biosynthesis of the polyketide segment of erythromycin.
CC eryA is organised in modules and each module takes care of one
CC condensation step. The precise succession of elongation steps is dictated
```

```
CC by the genetic order of the modules. This fragment may be specifically
CC altered such that novel polyketide molecules of desired structure are
CC produced. Three types of alteration may be produced; those inactivating a
CC single function in a module which does not arrest acyl chain growth;
CC those inactivating a single function in a module which does affect chain
CC growth; and those affecting an entire module. The mutations may be
CC introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN
CC field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;
Query Match 66.2%; Score 397.8; DE 2; Length 29879;
Best Local Similarity 78.9%; Pred. No. 1.1e-57;
Matches 474; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Q 1 GCGTCCTAAGCGGCGGGAACGCGCTTCTGACGCGCTTCCGCGCGGTGTCGCGAGT 60
D 24699 GCGTCCTAAGCGGCGGGAACGCGCTTCTGACGCGCTTCCGCGCGGTGTCGCGAGT 24758
Q 61 GCGCTGCGCGGTGACCTGATGCGCTGCGGTCTGTGGGCGCGGAGAACATGCGCGGTACC 120
D 24759 GCGCGCGCGGTGACCTGATGCGCTGCGGTCTGTGGGCGCGGAGAACATGCGCGGAGAC 24818
Q 121 GAGGGCGGCGACTAATCTGCGGACGCGGCGCTGCGGCGCTGCGGCGGCGGCGGCGATC 180
D 24819 GAGGGCGGCGACTAATCTGCGGACGCGGCGCTGCGGCGCTGCGGCGGCGGCGGCGATC 24878
Q 181 GAGGAGCTGGGAGCAACCTCGGACGCGGCGGAGACCGGCGGTGTCGCGGTGCGACTGAC 240
D 24879 GAGGAGCTGGGAGCAACCTCGGACGCGGCGGAGACCGGCGGTGTCGCGGTGCGACTGAC 24938
Q 241 CGGAGGCGGTGTCGCGAATGTTTCAACCGCGCGCGCGCGCGCGCGCTTTCGACGAATC 300
D 24939 CGGAGGCGGTGTCGCGAATGTTTCAACCGCGCGCGCGCGCGCGCGCTTTCGACGAATC 24998
Q 301 GGTGGGCTCGCGCGCGCGCGCGCGGAGAGACCGGTGAGGAATTCGATCTTCGCGCGGCTG 360
D 24999 GCGGCTCGCGCGCGCGCGCGCGGAGAGACCGGTGAGGAATTCGATCTTCGCGCGGCTG 25058
Q 361 GCGTCGATGCGCGGCGGAGCGGACGTCAGAGCATGTCGCGCGCGGTGTCGCGAGCGAGTG 420
D 25059 GCGGCTCGTCGACGCGCGGAGCGGAGCGGAGCATCTTCGCACTGATTCGCGCGGAGTTC 25118
Q 421 GCACGGGTGCTGGGCGCCACGGGACACCGGTGATCGAGGGTGAAGCTGCGCTTCCGCGAC 480
D 25119 GCGCGGTGCTGGGCGCCACGGGACACCGGTGATCGAGGGTGAAGCTGCGCTTCCGCGAC 25178
Q 481 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTGGCGGAAACCGGCTCGCGGCGGTGACCGGG 540
D 25179 CTGGGATTTCGACTCCATGACCGCGCGTGAACCTGGCGGAAACCGGCTCGCGGCGGTGACCGGG 25238
Q 541 GTCCGGGTGGCGACGACCATGCTTTTGACCAACCGGACAGTGAACCGGCTTCGCGCGGAC 600
D 25239 GTCCGGGTGGCGACGACCATGCTTTTGACCAACCGGACAGTGAACCGGCTTCGCGCGGAC 25298
Q 601 T 601
D 25299 T 25299

RESULT 4
ADM97245
ID ADM97245 standard; DNA; 9510 BP.
XX
AC ADM97245;
XX
DT 01-JUL-2004 (first entry)
XX
DE Synthetic polyketide synthase gene method related synthetic DB853 DNA.
KW ds; gene; polyketide synthase; PCR; medicine; agriculture;
KN synthetic gene; DB853.
XX
```









Db 4294 TGAAGGCGTTGCTGCTCGCCTTCACTCCGCGCGCCGACAGCCGCTGCTGACAGACTG 4353  
Qy 301 GGTGGGCTCG-----CGCCGGGCGGAGAGACCGGCTCAG 336  
Db 4354 CCCGAGGCGCGGAGGTCAATGACGCCACCGGACGAGGCGGAGAGACACCGGACGC 4413  
Qy 337 GAATCGATCTCGCCGCGGCTGAGCGTCAATGCGGAGGCGGAACGTCAAGCATATGTC 396  
Db 4414 GCGCCCGCGCTGCGCCAGCACTCAACGCGCGCCGAGGCGGAGAGAGAGGCACTGCTC 4473  
Qy 397 GCCCGCTGCTGCGAGCGGAGGTGCAAGCGGTCTGGGCGACGCGACCGGACGTATC 456  
Db 4474 CTGCAACTGTGCTCCGACCGCGCTGCGCGCGCTCTCGGCTACGCGGCGCCGACGCGATC 4533  
Qy 457 GAGGTGACGTCGCTTCCGTAACCTGAGATTCGATCCAGACCGCGCTGACGCTGCG 516  
Db 4534 GAGCGGCGCGCGCTTCAAGAGAGTGGCTTCACTCCGACCTCCGTAACCTGCGC 4593  
Qy 517 AACCGGCTCGCGGCGGTGACCGGAGGTCGAGGTGAGCGACCATGCTTTCACACCG 576  
Db 4594 AACCGCTGAACGCGGCGGCGGCTCAAGCTGCGCGCCACCTCTGCTTTCACACCG 4653  
Qy 577 ACAGTGACCGCGCTCAACCGCGCAC 600  
Db 4654 ACGCCCACTGCTCTGCGCGGAC 4677

RESULT 8  
AEB87003  
ID AEB87003 standard; DNA; 9684 BP.  
XX  
AC AEB87003;  
XX  
DT 06-OCT-2005 (first entry)  
XX  
DB Streptomyces albusensis polyketide synthase ORF16 DNA SEQ ID NO:34.  
XX  
KW polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;  
KM gene; de.  
XX  
OS Streptomyces albusensis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..9684  
FT /tag=a  
FT /product= "polyene polyketide synthase ORF16"  
XX  
CA2453071-A1.  
XX  
PD 03-APR-2004.  
XX  
PF 21-JAN-2004; 2004CA-02453071.  
XX  
PR 21-JAN-2003; 2003US-0441123P.  
XX  
PR 13-MAY-2003; 2003US-0469810P.  
XX  
PR 01-AUG-2003; 2003US-0491516P.  
XX  
PR 13-AUG-2003; 2003US-0494568P.  
XX  
PA (ECOP) ECOPIA BIOSCIENCES INC.  
XX  
PI Farnet CM, Zazopoulos E;  
XX  
DR WPI; 2004-348649/33.  
XX  
DR P-PSDB; AEB87002.  
XX  
PT Isolated, purified or enriched nucleic acid useful for producing  
XX  
XX polyketides, encodes domains of the polyketide synthase system.  
XX  
PS Claim 2; SEQ ID NO 34; 550pp; English.  
XX  
CC The invention relates to an isolated, purified or enriched nucleic acid  
XX which can be used for the production of a polyketide. The nucleic acid

CC encodes at least one domain of the polyketide synthase system formed by  
CC the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,  
CC AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)  
CC an isolated, purified or enriched nucleic acid for the production of a  
CC polyketide of formula I comprising a nucleic acid selected from the group  
CC consisting of AEB86865, AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,  
CC AEB86878, AEB86880 and AEB86882; (2) an isolated, purified or enriched  
CC nucleic acid for the production of a polyene polyketide; (3) an  
CC expression vector comprising the novel nucleic acid or the nucleic acid  
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)  
CC the cosmid deposited under INAC accession no. 250203-01, 250203-02,  
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for  
CC the production of a polyketide. The nucleic acid can be used for  
CC production of a polyketide, for use as a fungal and cancer cell growth  
CC inhibitors. The present sequence encodes Streptomyces albusensis polyene  
CC polyketide synthase ORF16, which is used in the exemplification of the  
CC present invention.

XX  
SQ Sequence 9684 BP; 1211 A; 3372 C; 3596 G; 1505 T; 0 U; 0 Other;

Query Match 34.5%; Score 207.6; DB 13; Length 9684;

Best Local Similarity 60.3%; Pred. No. 5,2e-26;

Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

Qy 1 GCTCTCAACGCGCGGCAACGCTTCTTCAACGCTTTCGCTGCTGCTGCGGCACT 60  
Db 4054 GGCACCTAACGCGCGGCAACGCGCTTCTTCAACGCTTTCGCTGCTGCTGCGGCACT 4113  
Qy 61 GGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
Db 4114 GAGCTGCGCGCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4173  
Qy 121 GAGGCGGCGCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
Db 4174 GCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4233  
Qy 181 GAGGAGTGGCGGCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
Db 4234 GCGGCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4293  
Qy 241 GCGGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
Db 4294 TGAAGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4353  
Qy 301 GGTGGGCTCG-----CGCCGGGCGGAGAGACCGGCTCAG 336  
Db 4354 CCCGAGGCGCGGAGGTCAATGACGCCACCGGACGAGGCGGAGAGACACCGGACGC 4413  
Qy 337 GAATCGATCTCGCCGCGGCTGAGCGTCAATGCGGAGGCGGAACGTCAAGCATATGTC 396  
Db 4414 GCGCCCGCGCTGCGCCAGCACTCAACGCGCGCCGAGGCGGAGAGAGAGGCACTGCTC 4473  
Qy 397 GCCCGGCTGCTGCGAGCGGAGGTGCAAGCGGTCTGGGCGGACCGGACCGGATGTC 456  
Db 4474 CTGCAACTGTGCTCCGACCGCGCTGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 4533  
Qy 457 GAGGTGACGTCGCTTCCGTAACCTGAGATTCGATCCAGACCGCGCTGACGCTGCG 516  
Db 4534 GAGCGGCGCGCGCTTCAAGAGAGTGGCTTCACTCCGACCTCCGTAACCTGCGC 4593  
Qy 517 AACCGGCTCGCGGCGGTGACCGGAGGTCGAGGTGAGCGACCATGCTTTCACACCG 576  
Db 4594 AACCGCTGAACGCGGCGGCGGCTCAAGCTGCGCGCCACCTCTGCTTTCACACCG 4653  
Qy 577 ACAGTGACCGCGCTCAACCGCGCAC 600  
Db 4654 ACGCCCACTGCTCTGCGCGGAC 4677

RESULT 9  
AEB86800  
ID AEB86800 standard; DNA; 9684 BP.  
XX



FT	/product= "Polyketide synthase ORF15 protein"
FT	94057. .103740
FT	/*tag= h
FT	/product= "Polyketide synthase ORF16 protein"
FT	103789. .126321
FT	/*tag= i
FT	/product= "Polyketide synthase ORF17 protein"
FT	126389. .138007
FT	/*tag= j
FT	/product= "Polyketide synthase ORF18 protein"
FT	complement(138063. .139079)
FT	/*tag= k
FT	/product= "Polyketide synthase related ORF19 protein"
FT	complement(139266. .140117)
FT	/*tag= l
FT	/product= "Polyketide synthase related ORF20 protein"
FT	complement(140114. .141103)
FT	/*tag= m
FT	/product= "Polyketide synthase related ORF21 protein"
FT	14183. .142436
FT	/*tag= n
FT	/product= "Polyketide synthase related ORF22 protein"
FT	142440. .143054
FT	/*tag= o
FT	/product= "Polyketide synthase related ORF23 protein"
FT	143133. .144119
FT	/*tag= p
FT	/product= "Polyketide synthase related ORF24 protein"
FT	144116. .145102
FT	/*tag= q
FT	/product= "Polyketide synthase related ORF25 protein"
FT	145099. .145743
FT	/*tag= x
FT	/product= "Polyketide synthase related ORF26 protein"
FT	145818. .147230
FT	/*tag= b
FT	/product= "Polyketide synthase related ORF27 protein"
FT	complement(147306. .148967)
FT	/*tag= t
FT	/product= "Polyketide synthase related ORF28 protein"
FT	complement(149176. .149871)
FT	/*tag= u
FT	/product= "Polyketide synthase related ORF29 protein"
FT	complement(149868. .150789)
FT	/*tag= v
FT	/product= "Polyketide synthase related ORF30 protein"
FT	complement(150769. .153765)
FT	/*tag= w
FT	/product= "Polyketide synthase related ORF31 protein"
FT	154485. .156041
FT	/*tag= x
FT	/product= "Polyketide synthase related ORF32 protein"
FT	156075. .157064
FT	/*tag= y
FT	/product= "Polyketide synthase related ORF33 protein"
FT	157308. .158873
FT	/*tag= z
FT	/product= "Polyketide synthase related ORF34 protein"
FT	158970. .160202
FT	/*tag= aa
FT	/product= "Polyketide synthase related ORF35 protein"
FT	160199. .161719
FT	/*tag= ab
FT	/product= "Polyketide synthase related ORF36 protein"
FT	161924. .162577
FT	/*tag= ac
FT	/product= "Polyketide synthase related ORF37 protein"
FT	162723. .164051
FT	/*tag= ad
FT	/product= "Polyketide synthase related ORF38 protein"
XX	
PN	WO2004065401-A1.

[illegible]

QY 517 AACCGCTCGCGGCGGTGACCGGGGTCGGGTCGACGACCATGCTTTCGACCCCG 576  
DB 98650 AACCGCTCGAACCGGCGCGACGCGGCTCAAGCTGCGCGCCACCTCGTTCGACCAACCG 98709  
QY 577 ACAGTGAACGCGCTCAACCGGCGAC 600  
DB 98710 ACGCCCAACGCTCTCGCCCGGAC 98733

## RESULT 11

ID AEB86862 standard; DNA; 164051 BP.

AC AEB86862;

DT 06-OCT-2005 (first entry)

DE Streptomyces aizumensis polyketide synthase contig 2 SEQ ID NO:18.

KM polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;  
de.

OS Streptomyces aizumensis.

PN CA2453071-A1.

PD 03-APR-2004.

PF 21-JAN-2004; 2004CA-02453071.

PR 21-JAN-2003; 2003US-0441123P.

PR 13-MAY-2003; 2003US-0469810P.

PR 01-AUG-2003; 2003US-0491516P.

PR 13-AUG-2003; 2003US-0494568P.

PA (ECOP-) ECOPIA BIOSCIENCES INC.

PI Farnet CM, Zazopoulos E;

DR WPI; 2004-348649/33.

PT Isolated, purified or enriched nucleic acid useful for producing  
XX polyketides, encodes domains of the polyketide synthase system.

PS Disclosure; SEQ ID NO 18; 550bp; English.

XX The invention relates to an isolated, purified or enriched nucleic acid  
CC which can be used for the production of a polyketide. The nucleic acid  
CC encodes at least one domain of the polyketide synthase system formed by  
CC the polyketide synthases of AEB86865, AEB86867, AEB86869, AEB86871,  
CC AEB86873, AEB86875, AEB86877, AEB86879 and AEB86881. Also described: (1)  
CC an isolated, purified or enriched nucleic acid for the production of a  
CC polyketide of Formula I comprising a nucleic acid selected from the group  
CC consisting of AEB86866, AEB86868, AEB86870, AEB86872, AEB86874, AEB86876,  
CC AEB86877, AEB86880 and AEB86882; (2) an isolated, purified or enriched  
CC nucleic acid for the production of a polyene; polyketide; (3) an  
CC expression vector comprising the novel nucleic acid or the nucleic acid  
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)  
CC the cosmid deposited under IDAC accession no. 250203-01, 250203-02,  
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for  
CC the production of a polyketide. The nucleic acid can be used for  
CC production of a polyketide, for use as a fungal and cancer cell growth  
CC inhibitors. The present sequence represents a Streptomyces aizumensis  
CC polyene polyketide synthase related contig 2 DNA sequence, which is used  
CC in the exemplification of the present invention.

SQ Sequence 164051 BP; 21253 A; 59641 C; 58659 G; 24298 T; 0 U; 0 Other;

Query Match 34.5%; Score 207.6; DB 13; Length 164051;  
Best Local Similarity 60.3%; Pred. No. 3.7e-26;  
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

QY 1 GCCTCTACCGGCGGCAACGCTCTCTTTCGACGCTTTCGCGGTGTCGTCGCGAGT 60  
DB 98110 GGCACTACGCGCGGGCAACGGGTTCTTGAAGCTTTCGACAGTACCGCGGCAATCC 98169  
QY 61 GGGCTGCGGTGACCTTCGATGCGCTTGGGCTTGTGGGCGGGCAACATGCGGTACC 120  
DB 98170 GGAATGCGGCGACCTTCGATGCGCTTGGGCTTGTGGGCGGGCAATGCGGAGGCG 98229  
QY 121 GAGGCGGCGACTACGTCGCGGCGACGAGGCTGCGGCGGCAATGACCGGCGGCGCATC 180  
DB 98230 GCGGTGATGACGGAATGCGCGCGCGGAGTCAAGATGTCGCGGAGCGCGTC 98289  
QY 181 GAGAGCTGCGGACCAACCTTGAACGCGGCGGACCCCTGAGTGTGTCGTCGTCGAC 240  
DB 98290 GCGGCACTCAGACACCGCTGACCGCGACGACACACCTTACCGTTCGCGCATGAG 98349  
QY 241 CGGAGCGGTTGCTGCAACTGTTACCGCGCGCGCGCGCGGCGGCTTTCGACGAACTC 300  
DB 98350 TGGAGCGCTTCTGCTCTGCGCTTCACTCGGCGCGGCGGCGGCTGTCGACGACTG 98409  
QY 301 GGTGGGCTCG-----CGCGGCGCGGAGAGACCGGTACG 336  
DB 98410 CCGAGGCGCGGAGGATCAAGCGCACGCGACGAGAGCGCGAGACACCGGACG 98469  
QY 337 GAATCGATCTCGCGCGCGGCTGAGTCGATCGGAGCGGACGTCAGACGATGTC 396  
DB 98470 GCGCGCGCTGCGCGGACGCTGACCGCGCGCGCGCGGCGGACGAGGCGACTGCTC 98529  
QY 397 GCGCGGCTGCTCGACCGGAGTGAAGCGGTGCGGCGGCGGCGGCGGCGGCGGATC 456  
DB 98530 CTCGAATGCTGCGGACCGCGCGGCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 98589  
QY 457 GAGCGTGAAGTGTGCTTCCGTGACCTGGGATTCGATTCGATGACCGCGCTGACCTGCGG 516  
DB 98590 GAGCGGCGCGGCTTCAAGAGCTGCGCTTCACTTCCTGCACTTCGCGGAACTGCGG 98649  
QY 517 AACCGCTCGCGGCGGTGACCGGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576  
DB 98650 AACCGCTGAACGCGGCGGCGGCGGCTTCAAGCTGCGGCGGCGGCGGCGGCGG 98709  
QY 577 ACAGTGAACGCGCTCAACCGGCGAC 600  
DB 98710 ACGCCCAACGCTCTCGCCCGGAC 98733

## RESULT 12

ID AEB86659 standard; DNA; 164051 BP.

AC AEB86659;

DT 06-OCT-2005 (first entry)

DE Streptomyces aizumensis polyketide synthase contig 2 SEQ ID NO:18.

KM polyene; polyketide; polyketide synthase; antifungal; cytostatic; cancer;  
de.

OS Streptomyces aizumensis.

PN CA2453080-A1.

PD 03-APR-2004.

PF 21-JAN-2004; 2004CA-02453080.

PR 21-JAN-2003; 2003US-0441123P.

PR 13-MAY-2003; 2003US-0469810P.

PR 01-AUG-2003; 2003US-0491516P.

PR 13-AUG-2003; 2003US-0494568P.  
PA (ECOP-) ECOPIA BIOSCIENCES INC.



```
PI Farnet CM, Bachmann BO, Mcalpine JB, Zazopoulos E;
XX WPI; 2004-348650/33.
DR
XX Polyene polyketide compounds are bacterial cell growth inhibitors useful
PT for treating e.g. fungal infections and cancer.
PS
XX Disclosure; SEQ ID NO 18; 144pp; English.
CC The invention relates to an isolated, purified or enriched nucleic acid
CC which can be used for the production of a polyketide. The nucleic acid
CC encodes at least one domain of the polyketide synthase system formed by
CC the polyketide synthases of ABB86662, ABB86664, ABB86666, ABB86668,
CC ABB86670, ABB86672, ABB86674, ABB86676 and ABB86678. Also described: (1)
CC an isolated, purified or enriched nucleic acid for the production of a
CC polyketide of Formula I comprising a nucleic acid selected from the group
CC consisting of ABB86663, ABB86665, ABB86667, ABB86669, ABB86671, ABB86673,
CC ABB86675, ABB86677 and ABB86679; (2) an isolated, purified or enriched
CC nucleic acid for the production of a polyene polyketide; (3) an
CC expression vector comprising the novel nucleic acid or the nucleic acid
CC of (1) or (2); (4) a host cell transformed with the vector of (3); (5)
CC the cosmid deposited under ID# accession no. 250203-01, 250203-02,
CC 250203-03, 250203-04, and 250203-05; and (6) an isolated polypeptide for
CC the production of a polyketide. The nucleic acid can be used for
CC production of a polyketide, for use as a fungal and cancer cell growth
CC inhibitors. The present sequence represents a Streptomyces albusensis
CC polyene polyketide synthase related contig 2 DNA sequence, which is used
CC in the exemplification of the present invention.
XX
SQ Sequence 164051 BP; 21253 A; 59641 C; 58859 G; 24298 T; 0 U; 0 Other;

Query Match 34.5%; Score 207.6; DB 13; Length 164051;
Best Local Similarity 60.3%; Pred. No. 3.7e-26;
Matches 376; Conservative 0; Mismatches 224; Indels 24; Gaps 1;

QY 1 GCCTCTACGCGCGCGGAGACGCTCTCTGACGCGCTTCGCGCGCTGCTGCGCGAGT 60
DB 98110 GGCACACTAGCGCGCGGAGACGCGCTCTCTGACGCGCTTCGCGCGAGTACCGCGCGAGTCC 98169
QY 61 GGGCTCGCGGATCACTCAATGCGCTGCGGAGTCTGCGCGCGGAGACATGCGCGGATC 120
DB 98170 GGACTGCGCGGATCACTCAATGCGCTGCGGAGTCTGCGCGCGGAGACATGCGCGGATC 98229
QY 121 GAGCGCGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGCGGAGTCTGCGGAGT 180
DB 98230 GCGGCTGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98289
QY 98230 GCGGCTGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98289
DB 181 GAGGAGCTGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 240
DB 98280 GCGGAGCTGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98349
QY 241 CGGAGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 300
DB 98350 TCGGAGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98409
QY 301 GGTGGGATCGG-----CGCGCGGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 336
DB 98410 CGCGAGCGGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98469
QY 337 GAATGCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 396
DB 98470 GCGCGCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98529
QY 397 GCGCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 456
DB 98530 CTGGAATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98589
QY 457 GAGCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 516
DB 98590 GAGCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98649
QY 517 AACCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 576

DB 98650 AACCGGATCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98709
QY 577 ACAGTGAGCGGCTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 600
DB 98710 ACAGTGAGCGGCTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 98733

RESULT 13
AAD55819
ID AAD55819 standard; DNA; 5355 BP.
XX
XX AAD55819;
AC
XX 27-OCT-2003 (revised)
DT 07-AUG-2003 (first entry)
XX
DB Micromonospora carbonacea polyketide synthase (PKS) type I gene #9.
XX
XX Macrolide; rosamycin; polyketide; polyketide synthase; PKS; enzyme;
KM gene; ds.
XX
XX Micromonospora carbonacea.
OS
XX
XX Key Location/Qualifiers
FH 1. .5355
FT CDS /*tag= a
FT /product= "Polyketide synthase"
XX
XX CA2391131-Al.
XX
XX 19-NOV-2002.
XX
XX 26-JUL-2002; 2002CA-02391131.
XX
XX 26-JUL-2001; 2001US-0307629P.
XX
XX (BCOP-) BCOPRIA BIOSCIENCES INC.
XX
XX Yang X, Staffa A, Farnet CM;
PI WPI; 2003-343556/33.
DR P-PSDB; AAE37003.
XX
XX Novel isolated polypeptide involved in biosynthesis of macrolides by
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,
PT preferably for biosynthesis of rosamycin.
XX
XX Claim 1; Page 176-176; 206pp; English.
XX
XX The invention relates to genes and proteins involved in the biosynthesis
XX of macrolides by microorganisms. In particular it relates to the nucleic
XX acids forming the biosynthetic locus for rosamycin (a 16-member
XX macrolide antibiotic) from Micromonospora carbonacea. The invention is
XX useful for the biosynthesis of macrolides by microorganisms. It allows
XX direct manipulation of macrolides and related chemical structures by
XX chemical engineering of the proteins involved in the biosynthesis of
XX rosamycin. It is useful to catalyze certain biochemical reactions, in
XX vitro or in vivo, to direct or enhance the synthesis or modification of a
XX polyketide, polyketide substrate or its precursor. The present sequence
XX is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-
XX OCT-2003 to standardise OS field)
XX
SQ Sequence 5355 BP; 516 A; 2056 C; 2107 G; 676 T; 0 U; 0 Other;

Query Match 34.0%; Score 204.4; DB 10; Length 5355;
Best Local Similarity 60.0%; Pred. No. 1.9e-25;
Matches 359; Conservative 0; Mismatches 236; Indels 3; Gaps 1;

QY 3 CTCTACGCGGCGGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 62
DB 3876 CGCTACGCGGCGGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 3935
QY 63 GCTGCGGATCACTCAATGCGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGTCTGCGGAGT 122
```







QY	QY	QY	QY
1	61	51167	51226
CCGCGCTCCGCGAGAGGCTACCCGGGGGGGTGAGGTGCATCTGGCTTACCGGACAGTGGAGT	AGGAGAGCGCCGCTGCACCTTCGCGGTCTACCGCTTCCAGACAGACAGAACTTCTGGCTCCCG	CCGCGCTTGGCGCGAGGGTACGACACAGCGGTGAGGTGCATCTGGCTGCGCGCTGTTCCCG	GGCGCGCGCGCGGTGATCTGGCGACAGCTTACCGCTTCCAGGCGGAGCGCTTACTGGCTGGAGC
60	120	51226	51286



APPLICANT: Eteghalian, Alireza  
TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
FILE REFERENCE: 56462007901  
CURRENT APPLICATION NUMBER: US/10/517, 939  
CURRENT FILING DATE: 2004-12-13  
PRIOR APPLICATION NUMBER: PCT/US03/19153  
PRIOR FILING DATE: 2003-06-16  
PRIOR APPLICATION NUMBER: 60/389,299  
PRIOR FILING DATE: 2002-06-14  
NUMBER OF SEQ ID NOS: 380  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 359  
LENGTH: 2724  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample.  
US-10-517-939-359

Query Match 10.5%; Score 63; DB 8; Length 2724;  
Best Local Similarity 45.5%; Pred. No. 4.4e-05;

Matches 285; Conservative 0; Mismatches 315; Indels 3; Gaps 1;

QY 14 GGCGTACACCCGGGGGTGGAGGTGCACTGGCGTACCGCAGTGGGTGAGGAGCGCCGGT 73  
DB 2046 GGCGTACACCGGGCCATGACCGCCGACCTGAGACGACCGGTGCGGCTACGTT 2105  
QY 74 CGACCTGCGGGTCTACCTGTTCCAGACGAACTTCTGCTCCGGTCCCTGGGCG 133  
DB 2106 CGCGAGGACATGCGCTCGACGAGGCGCGCTGACGAGCCGACCTGGGAGCGTGGCC 2165  
QY 134 GGTCCTCCGACACCGGGGAGAGTGGGCTTACGAGTCTGCTGGGACCCCGTCCAGCTCGG 193  
DB 2166 GCTGACACGAGTGAAGGGGCGGCGAGTTCCAGCTCCGCTGGGCGGCGACCACTGAC 2225  
QY 194 GCGGTCTCTGCTGCGGAGCGGGTCTGAGTGAACGAGCGGAGTACCCCGGCTG 253  
DB 2226 GGTGTTCTGCTCACTGACGAGCGGTGACGAGTCTGCTGCGGAGCGAGACTTA 2285  
QY 254 GACGAGCTGCTGCTGCGGAGCGGCTGGAACAGCGGGGCGACCGTCTGTTGACCGC 313  
DB 2286 CACGGTCTGCTGCGAGCGGCGAGGCGGAGCTGAGACGCGGTCAACCGGCGGAGGAGAACG 2345  
QY 314 GCAAGTGGCGCGCCCGGATGCGCGCCGCACTGACCGCG--TGAAGCGACCGCGCTGTC 370  
DB 2346 CTCCTGAGACCGCTGCTGCTGCGCTGCGCTGACGCGGAGCAAGCGGCGGACCTGTT 2405  
QY 371 CACTGTGCTCTCTGCTGCGCGCTGCGCGGAGGCGGAGTGTGACGACCCCGAGCTGGA 430  
DB 2406 CGACTTCGGATCATGACGAGCGCCGACACCTCCGGGTGGAAGTGTGAGGTCTGCGG 2465  
QY 431 CACCTTCGGGTGCTGCAAGCGCTGCGCGGAGCGGAGTGAAGTCCCTGCTGAGTGTG 490  
DB 2466 CACCTGACCTGCTGCTGAGGAGCTGCTTCTGCTGAGGTGTGAGGCGGCGGACCGGCC 2525  
QY 491 GACCAAGGAGCGCGCGCGCTGACCGTGGAGACGAGTGTGATCCGCGCCAGCGCATGTG 550  
DB 2526 GACCATTCAGCGGAGATGACCGCCGCTGAGGAGAGCGCAAGTGTGTCAACGACGAT 2585  
QY 551 CGGTGGGCTGCGCGGCGGTGAGGCGTGAAGTCCCGCGCGG 593  
DB 2586 CGGTATGAGAGCGGCTGCTGACGCGCGGAGGCGGAGATCCGG 2628

#### RESULT 5

US-11-052-554A-519  
Sequence 519, Application US/11052554A  
Publication No. US2005028866A1  
GENERAL INFORMATION:  
APPLICANT: Sachdeva, et al.  
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

FILE REFERENCE: 30853/40359A  
CURRENT APPLICATION NUMBER: US/11/052,554A  
CURRENT FILING DATE: 2005-02-07  
PRIOR APPLICATION NUMBER: US 60/589,227  
PRIOR FILING DATE: 2004-07-20  
PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
PRIOR FILING DATE: 2004-02-06  
NUMBER OF SEQ ID NOS: 763  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 519  
LENGTH: 5706  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-519

Query Match 10.3%; Score 62; DB 14; Length 5706;  
Best Local Similarity 49.0%; Pred. No. 6.3e-05;  
Matches 193; Conservative 0; Mismatches 200; Indels 1; Gaps 1;

QY 183 GTGACCTCGGGGCGCTCTCCCTGGGCGGACCGGTCTGTGTGACCGGAGCGGAGTA 242  
DB 2066 GTGACGTGCGGCTGCGGCGGCTGCGAGCAGCGCTACCGGTGCGCGGCTTCCGCGG 2125  
QY 243 CCCCCGCTGAGAGGAGCTGTGCTCGGACGCGCTTGAACAGCGCGGCGGACCGTCTG 302  
DB 2126 GCGCGCGGAGAGAG-CCGAGCGGCGGCGGACAGCGGTGTGGCGGCAACAGGCTCC 2184  
QY 303 TTGTGACCGCGGAGTGCAGCGCGCGGATGCGCGGCACTGACGCGCTGACGCGGACCC 352  
DB 2185 GCGGCGCGCGGCGGTGCAAGCGGCGGCAAGGCGGCGGAGTGCAGCGGCTCCGCGCG 2244  
QY 363 GCGCTGTCACTGTGTCTCTGCTGCGGCTGTGCGGAGGCGGCTGTGACAGACCC 422  
DB 2245 GACAACTCCAGCGGTGTGTTTCTGCGGCGGCGGCGGCGGAGTGTGCGGCGGCG 2304  
QY 423 AGCTGAGACACCTCGGCTGCTGCAAGGCGCTGCGGCGGCGGAGTGTGCGGCGGCG 482  
DB 2305 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2364  
QY 483 TGGCTGTGACGAGGAGCGCGCGCGCTGAGCGGCTGTGAGCGAGCGAGTGTGCGGCGG 542  
DB 2365 GGTAGTGCAGGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2424  
QY 543 GCCATGTGTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576  
DB 2425 GGTCTGCGGCTGCGGCGGCTGCGGCGGCTTGTGAGGCGG 2458

#### RESULT 6

US-11-096-568A-12512  
Sequence 12512, Application US/11096568A  
Publication No. US20060048240A1  
GENERAL INFORMATION:  
APPLICANT: Alexandrov, Nikolai et al.  
TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptide  
FILE REFERENCE: 2750-1592PUS2  
CURRENT APPLICATION NUMBER: US/11/096,568A  
CURRENT FILING DATE: 2005-04-01  
NUMBER OF SEQ ID NOS: 34471  
SEQ ID NO 12512  
LENGTH: 1450  
TYPE: DNA  
ORGANISM: Trifolium aestivum  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(1450)  
OTHER INFORMATION: Ceres Seq. ID no. 14302059  
US-11-096-568A-12512

Query Match 10.2%; Score 61; DB 11; Length 1450;  
Best Local Similarity 45.0%; Pred. No. 0.00014;  
Matches 229; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

```

QY      1 CGGCGCTGGCCGAGGGGTTCACCCGGGGGGTGGAGGTTCAGCTGCGCTACCGCAGTGGTG 60
Db      59 CGCGCGCGCGCGCGCGCGCGCGAGCCCTCTCCGCGCAGACCGCGCGCTACGTGTGGACT 118
QY      61 AGGAGCGCCCGGTTCGACCTGCGCGGTTCACCCGTTCCAGACGACGAACTTCTGCTCCCG 120
Db      119 AGCGCGCTGCGCGCCAGGAAGACAGAAACAGCTTACTAAGCCCTGCTCATCTCCCGCGG 178
QY      121 TCCCGCTGGGCCGGGGTTCGCCGACACCGGGGACGAGTGGCGTTACCACTTGCTTGGCACC 180
Db      179 CGGCGCGCGCGGTGACCTGCTGCCGCTCCGCTGACGACGCGCGCGCCCTTCGCGACACG 238
QY      181 CGGTGCACCTCGGCGCGGTCTCTCCCTGACGAGGAGGCTCTGTTGTTGACCGGACGCGCAG 240
Db      239 CTTTGACCTTCGTATTCACACAAAGCTTACGCGCACAACATGGCGCGCCCAAGTCCAGGCT 298
QY      241 TACCCCGCGCTTGGACGAGCGAGTGGTCCGCGACGCGCTTGAAACAGCGCGGGGCGACCTG 300
Db      299 TCTCGCGCGGCTACCCCTCCGCTCCGCTGGTTCGACCCGCGCCACGCGCATGACCGCTCC 358
QY      301 TGTGTGACACCGGCGAGTGGCGCGCCCGGATCGGCGCGCGACCTTCGACGCGCTGACGCG 366
Db      359 ACAACGCGACCTTCACATGCTCCAGGTGCTTCGAGCTCGACGCTCCGCTACCGACGCG 418
QY      361 CGGCGCTGTTCACATGTGTGCTCTCTGTGCTGCGCTCGCGCGAGGCGGGGTGTGTCAGACAC 420
Db      419 CTTCCGCGCTTCGCTGACGCGGAGGACACCGCCACACTTCGGCATTCCCAGACGAGTGG 478
QY      421 CCAGCTTGACACCTTCGCGGTGTGCTCAGGCGCTCGGCGGACGCGGAGTGCAGCTCCCG 480
Db      479 TCGTTCACGACGCGCGCGCTGCGGACTCGGAGCTCTCGCGCGCGCTCCGCTTCCGCG 538
QY      481 TGTGCTGTGACACGAGGACGCGCGCGCC 509
Db      539 TCATGCGCAAGCCCTCTGCTGCGGACGCGC 567

RESULT 7
US-11-205-109-1
/ Sequence 1, Application US/11205109
/ Publication No. US20050287641A1
/ GENERAL INFORMATION:
/ APPLICANT: Parmet, Chris
/ APPLICANT: Zazopoulos, Emmanuel
/ APPLICANT: Staefa, Alfredo
/ TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANTIN BIOSYNTHESIS
/ FILE REFERENCE: 3002-205
/ CURRENT APPLICATION NUMBER: US/11/205,109
/ CURRENT FILING DATE: 2005-08-17
/ PRIOR APPLICATION NUMBER: US/09/976,059
/ PRIOR FILING DATE: 2001-10-15
/ PRIOR APPLICATION NUMBER: US 60/259,924
/ PRIOR FILING DATE: 2000-10-13
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 1
/ LENGTH: 88421
/ TYPE: DNA
/ ORGANISM: Actinoplanes sp.
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (3118)..(4032)
/ OTHER INFORMATION: ORF 2; positive strandedness
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (2077)..(3078)
/ OTHER INFORMATION: ORF 1; positive strandedness
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (4038)..(5048)
/ OTHER INFORMATION: ORF 3; positive strandedness
/ FEATURE:

```

NAME/KEY: misc_feature	LOCATION: (6665)..(5814)	OTHER INFORMATION: ORF 4, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (7703)..(6673)	OTHER INFORMATION: ORF 5, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (9464)..(8110)	OTHER INFORMATION: ORF 6, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (9691)..(10761)	OTHER INFORMATION: ORF 7, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (12751)..(10829)	OTHER INFORMATION: ORF 8, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (13617)..(12802)	OTHER INFORMATION: ORF 9, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (15203)..(13614)	OTHER INFORMATION: ORF 10, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (15591)..(15863)	OTHER INFORMATION: ORF 11, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (15880)..(13035)	OTHER INFORMATION: ORF 12, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (19032)..(39713)	OTHER INFORMATION: ORF 13, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (39713)..(65800)	OTHER INFORMATION: ORF 14, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (65826)..(66310)	OTHER INFORMATION: ORF 15, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (66546)..(67370)	OTHER INFORMATION: ORF 16, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (67384)..(70059)	OTHER INFORMATION: ORF 17, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (70059)..(70662)	OTHER INFORMATION: ORF 18, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (70659)..(71906)	OTHER INFORMATION: ORF 19, positive strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (77439)..(71964)	OTHER INFORMATION: ORF 20, negative strandedness
FEATURE:		
NAME/KEY: misc_feature	LOCATION: (74216)..(73563)	OTHER INFORMATION: ORF 21, negative strandedness
FEATURE:		
NAME/KEY: misc_feature		

```
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87454)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1

Query Match 10.1%; Score 60.4; DB 14; Length 88421;
Best Local Similarity 49.1%; Pred. No. 7.8e-05;
Matches 160; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 191 CGGCGGCTCTCCCTGCGCGGACGAGGCTCTGTGTGACCGGAGCGGAGTACCCCGGC 250
Db 16476 CGGCGTGTGGGGGCTGCGCGGAGCTGCTCGGGGCGGGGCGGCGCGGCGTGC 16535
QY 251 CTGGAAGGAGAGTGTCTCGGACGCGGCTGGAACGCCGGGGGCGACCGTGTGTGAC 310
Db 16536 CGGACGCGCGGCTGCGGACGAGCAGCCTCTCCGCGGAGCGGCGGCGGCTGCGCGC 16595
QY 311 CGGCGAGTGGCGGCGCGGAGTGGGCGCGGACCTGACGCGGTCGAGCGGACCGGCGCTGC 370
Db 16596 GGGGGGCGGGGCGGCTGCGGACGAGCTGTGGCGGCGGCGGATGCGGCGGCGCTCTA 16655
QY 371 CACTGTGTCTCTCTGCTGCGGCTCGCGGAGGGGCGGTGTGACGACCCGAGCTTGA 430
Db 16656 CCAAGCACCGGCTCACCGGCGCGCGGAGCGTGTGTGGGCGTCCGCGCGCGGCGAC 16715
QY 431 CACCTCGGCTGTGTGACGCGCTCGGCGGACCGGAGTGTGACGCTCCCTGTGTGCTGGT 490
Db 16716 CACCGCGCGGAGTGTGAGCGTGCCTCGGCGATGAGTGTGCTCCGCTGCGCTCGC 16775
QY 491 GACCAAGGAGCGCGCGCGCTGACCG 516
```

```
Db 16776 GGTACGCGCGGCGGACGCGTGGCGC 16801
13
RESULT 8
US-11-096-568A-8975
Sequence 8975, Application US/11096568A
Publication No. US20060048240A1
GENERAL INFORMATION:
APPLICANT: Alexandrov, Nikolai et al.
TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
FILE REFERENCE: 2750-1592PUS2
CURRENT APPLICATION NUMBER: US/11/096,568A
NUMBER OF SEQ ID NOS: 34471
SEQ ID NO 8975
LENGTH: 1328
TYPE: DNA
ORGANISM: Triticum aestivum
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(1328)
OTHER INFORMATION: Ceres Seq. ID no. 13596902
US-11-096-568A-8975

Query Match 10.0%; Score 60.2; DB 11; Length 1328;
Best Local Similarity 47.9%; Pred. No. 0.00021;
Matches 173; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 104 GAACCTGTGCTCCCGGTCCTCCCTGGGCGGCGGTCGCGACACCGGAGAGTGGGCTTA 163
Db 169 GCACTCTCGCGGCGGCACTCTCCGTGCGGACGTCGACCTGCGGAGTGAACAATGACTA 228
QY 164 CCACTGCTGTGACACCCCGTGTGACCTGCGGCGGTCCTCCCTGCGGAGCGGTCGTGT 223
Db 229 CTCGCGGCGCGCGCGCGTGTGCGGCGGAGCGCTCTCTGCTGCTGCGGACCGCGGCTCGA 288
QY 224 GGTGACCGGAGCGGCACTACCCCGGCTGTGACGAGACGTGTGCGGAGGCTGTGAACA 283
Db 289 GCCCGCGCGCTGTGTGTGTGCGGACCGGCGCGGCTGTGACCATTTGCGCAACAATGA 348
QY 284 GCGCGGGGCGAGCGTGTGTGTGACCGGAGTGTGCGGCGCGGAGTGTGCGGCGGCTACT 343
Db 349 CCGCGGCTGTGACGCTACCACTGACGCTCCGCGGCGGAGCGGCTGTGAT 408
QY 344 CCAAGCGCTGTGACGCGGCGGCTGTGACCTGTGTGCTCTGCTGCGGCTGCGGAGG 403
Db 409 CAAGCTGTGACAGTCTGTGCTCTCTGCGGAGTGTGACCGGCGGAGGCGGCGGCTGT 468
QY 404 CGGTGTGTGACGACCCGAGCGCTGTGACACCTTGTGCTGTGTGTCAGGCGCTGTGCGAGC 463
Db 469 CATGCGCGAGCGCTGTGCTGTGACACCCCTTGTGCGGCGGCGGCTGCGGCGGAGCGC 528
QY 464 C 464
Db 529 C 529

RESULT 9
US-11-052-554A-543/C
Sequence 543, Application US/11052554A
Publication No. US2005028886A1
GENERAL INFORMATION:
APPLICANT: Sachdeva, et al.
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
FILE REFERENCE: 30853/40359A
CURRENT APPLICATION NUMBER: US/11/052,554A
PRIOR FILING DATE: 2005-02-07
PRIOR APPLICATION NUMBER: US 60/589,227
PRIOR FILING DATE: 2004-07-20
PRIOR APPLICATION NUMBER: IN 173/DL/2004
```



```

: PRIOR FILING DATE: 2004-02-06
: NUMBER OF SEQ ID NOS: 763
: SOFTWARE: Patentin version 3.3
: SEQ ID NO 543
: LENGTH: 2514
: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-543

```

Query Match	10.0%	Score 59.8	DB 14	Length 2514
Best Local Similarity	47.9%	Pred. No. 0.00022		
Matches 172; Conservative	0	Mismatches 187	Indels 0	Gaps 0

Oy	17	CGCGTCGACACCGGTGGAACCTCGGGGGGTCTCCCTGAGCGGAGCGGGTCTGTGTGTATAC	22
Db	733	CGCGCGCGCGCGCGGTATCCGAGCTCGGTATGCCCCCGCGGCGCACCGAACCCCGCGGTGAG	67
Oy	230	CGAGGCGGCGAGTACCCCGCGGCTGTGAGCGGACGTGTGTCCGCGGACCGGCGCTTGAAACAGCGCGG	28
Db	673	CGAACAGGCGCGCGCTTTTCCGCGCGGACCCCGCGGCGCGCGCGGTGTGGCGCCGAAAGCGGCTTG	61
Oy	290	GGCGACCGTCTGTGTGTGTGACACCGCGGACGTGCGCGGCGCGGATCTGGCGCGCGCACTGACGCG	34
Db	613	CGATTCGCGCGCGGAGACCGCGCACCGCCCGCGCGCGCCCGAAACAGCAATGCGCGGCGGCGCGCGCGG	55
Oy	350	CGTGCACGGACCGCGCGCTGTTCACACTGTGTCTCTCTGTGCTTTCGCGCTCGCGGAGGGCGGTGC	40
Db	553	CGCGCGCGGCGCGCGCGGTCCCGTGTGGCGCTTCCCGCGCGGCGCGCGCGCGCGCGCGCGT	49
Oy	410	TGTGCAGGACCCCGACCTTGACACACCTTCGCGTTGTGTTCAGGCGGCTTGCGCGACCGGGAT	46
Db	493	TGCGGATGAGCCCGCCCGCGCGCGCGCGCTGTGCGCGCGCGCCCGCGGTGTGAAGCGCGCGCGCGC	43
Oy	470	CGAGGTCCCCCTGTGTGCGGTGTGACGAGGAGCGCGCGCGCGCGCGGTGACCGTTCGAGAGCGAG	528
Db	433	CGGATCCTTCGCGCGCGCGCGCTGTGTCCGATTACGACGCGCGCGGTCCCCCGCATTTGGCCCCG	375

**RESULT 10**

```

US-11-052-554A-546
; Sequence 546, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 546
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-546

```

	Query Match	Match Similarity	Score	59.4%	DB 14%	length	1632;
	Best Local	Similarity	49.6%	Pred. No.	0.003		
	Matches	180;	Conservative	0;	Mismatches	181;	Indels
						2;	Gaps
						1;	
Qy	207	GGCGAAGCGGCTCTCTGTGTATACCGAAGCGAGATACCC--CGGCTTGACAGCAACATCGTGG	264				
Db	938	GCTTGAGCGGTATACGTGTGCAACGGTGGCGCGGTATCCCGCAATCGGATTCACAGCGCGGTGG	997				
Qy	265	TCGCGACAGCGGCTTGAAACAGCGCGGAGCGACCGTCTGTGTGTGTGACACCGGCGAGTGCAGCG	324				
Db	998	ACGGTGGACGCGCGCGGTGACAGCAACGCGCTTATCGGTGATCGGATCCAGAGGCGGCTTCGCGCG	1057				

QY	325	CCCGAATCGGCGCGCACTCGACGCCGTTCGACCGGACCGCCCTGTCCACTGTGATCTC	384
Db	1058	GCCTCGGCGGGTTTCGGGAAAGTACCGGCGGCGACCGCGCGCTGTGGGGGCGCGCGCGCTG	1117
QY	385	TGCTCGGCGCTCGCCGAGGCGGCTGTGTGACGACCCCGACCTGTGACACCTTCGCTTTGG	444
Db	1118	TGCTCATTCGGACCGGCGGTTCGGCGGTTTTCGCGCGCCTTCGGTGGCGGCTTCGAACGGACCG	1177
QY	445	TCCAGGCGCTTCGCGCGCACCGGATTCGACGTTCCCTCGTGGTGGTGAACCAAGGAGACGCG	504
Db	1178	GGGGCGCGGCGCGCGCGCGCGCGCGCACCGGCGCCGACGCTGATTCGGGCTTCGGGCGCGGCGCG	1237
QY	505	CCGCCGTGACCGTTCGAGACGACGTCGATCCCGGCCAGGCGCATGTTCGTTGGGCTCGAGCC	564
Db	1238	GCGGCGGCGATTCGGCGGGTTTCGCGCTCAACGTGGGCAACCGCGCTTCGGCGGCTTCGGCGGCGC	1297
QY	565	GGG GGG 567	
Db	1298	AGG 1300	

## RESULT 11

```

US-11-254-686-8
; Sequence 8, Application US/11256686
; Publication No. US20060040877A1
; GENERAL INFORMATION:
; APPLICANT: Martin, Christine J.
; APPLICANT: Burns, Lesley S.
; APPLICANT: Graupner, Paul R.
; APPLICANT: Lewer, Paul
; APPLICANT: Vonsden, William A.
; APPLICANT: Wilkinson, Barrie
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Novel Spinosyn-Producing polyketide Synthases
; FILE REFERENCE: 62069
; CURRENT APPLICATION NUMBER: US/11/254,686
; CURRENT FILING DATE: 2005-10-20
; PRIOR APPLICATION NUMBER: US 60/358075
; PRIOR FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Saccharopolyspora erythraea
US-11-254-686-8

```

Query Match	9.9%;	Score 59.4;	DB 11;	Length 1685;
Best Local Similarity	47.0%;	Pred. No. 0.0003;		
Matches 254;	Conservative 0;	Mismatches 281;	Indels 6;	Gaps 2;

QY 63 GAGAGCCCGGGGTGACCTGGCCGGGTCTAACCCGTTCCAAAGACAGAACTTTGGGCTCCGGGCTG 122

Db 1000 GAGAGTGGGTGACACCGGGAGCTGGTGGCCACTGACCTGGCCGCCCGAGCTTCCGGCTG 1059

QY 123 CCCCTGGAGCCGGAGTCCCGGACACCGGAGCAGAGTGGCGTTTACCACTTCCTGGACCCC 182

Db 1060 CCGGTGGCCTTCAAGAGAGGAGATCCGGTCCGCCCTTGGAGAGTCCGGTCCCGGACGTTTCGTC 1119

QY 183 GTCGACCTGGAGCCGGTCTCCCTGGCCGGACCGAGTCTTGGTGGTGAACCGAGCCGACGTA 242

Db 1120 GAAAGGAGACCCGACCCCGAGTCTGGCCGCGCCGCTTCCAGGACGAGCGTTGACACGCGAGAGGC 1179

QY 243 CCCCCGGGCTGAGAGGAGCGTGGTCCGGAGCGGCGCTTGGACAGCGCGGGAGCAGCGTGGTG 302

Db 1180 TCTCTGGGCGGGAGTGTCCGACGCTGCAACCGGGGACGAGCGGACATGGCGGGTTCTTG 1239

QY 303 TTGTGCAACCGGACAGTCCGCGGCGCCCGAGATCGGACCGGACCTG--ACGCGTTGACGGC 359

Db 1240 CTGGGCGGGGCCAAGCGCTTCAACCGGCGGGCTGGCCCGTCACTGAGACCGCGCGCTTACGAC 1289

QY 360 ACCGCGCTGTCCACTGGTCTTCTTGGTCCGCGCTCCGCGAGGGCGAGTGGTGTGACGAC 419

Db 1300 GACGTGAGGCGCCGACCCGCTCTCTGCGCGAGTTCCGCGCGCCGCGGACGAGAGACGAG 1359  
QY 420 CCCAGCTGAGACACCTCCGCTGTGTGTCAGAGCGCTCGCGCCACCCGAGATCGATCCCC 419  
Db 1360 CCGGCGAGTCCGCGCGTCACTG---AACGCGCACCCGACCTGTCTGCGAGCGGCTG 1416  
QY 480 CTGTGCTGTGTGACCAAGGAGCGCGCCGTCAGCTCGAGACGATCGATCCGCGC 539  
Db 1417 CTGCGGTGTGTCAACCGGACGACCGCGCTTGGCGGCGCGGACCGACGCGCGAGGCC 1476  
QY 540 CAGGCCATGATCGGTGCGCTCGGCGCGGCTGTGAGTCCCGCGCGCGGTGAGT 599  
Db 1477 ACCTTCGCGAGCTGTGAGGCTGAGCTGTGCTGAGCGCGGACGCTGCGCGCAAGTGTAGC 1536  
QY 600 G 600  
Db 1537 G 1537

## RESULT 12

US-11-052-554A-522/C  
; Sequence 522, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 522  
; LENGTH: 4146  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-522

Query Match 9.8%; Score 58.6; DB 14; Length 4146;  
Best Local Similarity 46.6%; Pred. No. 0.00037;  
Matches 224; Conservative 0; Mismatches 254; Indels 3; Gaps 1;

QY 120 GTCCCTCTGAGCGGCTCCCGACACCGGCGAGAGTGGGCTTACAGAGCTCGCTGGCAC 179  
Db 2046 GGGCGCGGCGGCGCGCGCTCAAGCGCGCTTGGCGACCGGCGCGCGCTGAGCGGCTTC 1987  
QY 180 CCGGTGACCTCGGCGGCTCTCTGCGCGGAGCGGAGTCTGTGTGACCGGAGCGGCA 239  
Db 1986 GCGGTCAACCCCAAGTTCGCGTGTGCGGAGCGGAGCTTCCCGCGCGCGCTTGGCC 1927  
QY 240 GATACCCCGGCTCGAGACGAGTGTGCGGAGCGGCTGGAACAGCGCGGCGGAGACCTTC 299  
Db 1926 GCGCGGCGGCGCTCGCGCGGAGCGGCGGAGTGTGCTGTGCGGAGTGTGCGGCTTC 1867  
QY 300 GTATTGTGACCGCGAGTGTGCGGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAG 359  
Db 1866 GCGCGGCGGCGCTCGCGCGGAGCGGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAG 1807  
QY 360 ACCGCGCTGTGCACTGTGTGTCTGTGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTG 419  
Db 1806 TGGGAGATGT 1750  
QY 420 CCCAGCTGAGACACCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 479  
Db 1749 GCGCGCTGT 1690  
QY 480 CTGTGCTGT 539  
Db 1689 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1630

QY 540 CAGGCCATGATCGGTGCGCTCGGCGCGGCTGTGAGTCCCGCGCGGTGAGT 599  
Db 1629 CCGCGCGGCGGCTGTGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1570  
QY 600 G 600  
Db 1569 G 1569

## RESULT 13

US-11-228-659-45  
; Sequence 45, Application US/11228659  
; Publication No. US20060064786A1  
; GENERAL INFORMATION:  
; APPLICANT: Brugiere, Norbert  
; TITLE OF INVENTION: Isopentenyl Transferase Sequences and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 1507  
; CURRENT APPLICATION NUMBER: US/11/228,659  
; PRIOR FILING DATE: 2005-09-16  
; PRIOR APPLICATION NUMBER: 60/610,656  
; PRIOR FILING DATE: 2004-09-17  
; PRIOR APPLICATION NUMBER: 60/637,230  
; PRIOR FILING DATE: 2004-12-17  
; PRIOR APPLICATION NUMBER: 60/696,405  
; PRIOR FILING DATE: 2005-07-01  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 984  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: OsIPT2 coding sequence (018830\_1)  
; NAME/KEY: CDS  
; LOCATION: (1)...(984)  
US-11-228-659-45

Query Match 9.7%; Score 58.2; DB 11; Length 984;  
Best Local Similarity 47.0%; Pred. No. 0.00061;  
Matches 214; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

QY 107 CTTCGCTCTCCGCGTCCCGCTGCGGAGTCCCGACACCGGCGAGAGTGTGAGTACCA 166  
Db 216 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 275  
QY 167 GCTCGCTGAGACCCCGTGTGAGCTGTGCGGCGGCTTCTGCGGCGGAGCGGAGTCTGTGT 226  
Db 276 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 335  
QY 227 GACCGGAGCGGAGTACCCCGGCTGTGAGCGAGTGTGTGCGGAGCGGCTGTGAAACAGCG 286  
Db 336 CCGCGGCGGCTGTGAACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 395  
QY 287 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 346  
Db 396 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 455  
QY 347 CCGCGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 406  
Db 456 CAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 512  
QY 407 TGT 466  
Db 513 CCGGCTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 572  
QY 467 GATCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 526  
Db 573 GATCGAGCGGCTTGT 632

Qy 527 CGTCGATCCGGCCAGGCGCAGTGGTGGGCTCG 561  
| | | | | | | | | | | | | | | | | | | |  
Db 633 CCTCCTCGCCGCGCATGCTCGAGAGCCGCGCTGCGCG 667

## RESULT 14

```

US-11-228-659-44
/ Sequence 44, Application US/11228659
/ Publication No. US20060064786A1
/ GENERAL INFORMATION:
/ APPLICANT: Brugiere, Norbert
/ TITLE OF INVENTION: Isopentenyl Transferase Sequences and
/ FILE OF INVENTION: Method of Use
/ FILE REFERENCE: 1507
/ CURRENT APPLICATION NUMBER: US/11/228,659
/ PRIOR FILING DATE: 2005-09-16
/ PRIOR APPLICATION NUMBER: 60/610,656
/ PRIOR FILING DATE: 2004-09-17
/ PRIOR APPLICATION NUMBER: 60/637,230
/ PRIOR FILING DATE: 2004-12-17
/ PRIOR APPLICATION NUMBER: 60/696,405
/ PRIOR FILING DATE: 2005-07-01
/ NUMBER OF SEQ ID NOS: 87
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 44

```

```

?      LENGTH: 4470
?      TYPE: DNA
?      ORGANISM: Oryza sativa
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: (1484) ... (2470)
?      FEATURE:
?      NAME/KEY: misc_feature
?      LOCATION: (0) ... (0)
?      OTHER INFORMATION: OsIRT2 full length sequence (018830_1
?

```

[illegible]

```

/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 2854, 2955, 2956, 2957, 2958, 2959, 2960, 2961, 2962, 2963,
/ LOCATION: 2964, 2965, 2966, 2967, 2968, 2969, 2970, 2971, 2972, 2973,
/ LOCATION: 2974, 2975
/ OTHER INFORMATION: n = A,T,C or G
US-11-228-659-44

```

Query Match	9.7%	Score 58.2;	DB 11;	Length 4470;
Best Local Similarity	47.0%	Pred. No. 0.00044;		
Matches 214; Conservative	0;	Mismatches 238;	Indels 3;	Gaps 1;

[illegible]

Qy	167	GCTGGCTGGACACCCGCTCAACTGGGGCGGTCTCTCTGGCCGGACGGGGCTGTGTGT	226
Db	1759	CCGCCCGAGGCGGGCGCGCGCGCTCCCGCGTCAAGCGGGCCGGGCGCCCTCTCCCGTCGT	1818

Qy	227	GACCGAGCGGCAGTACCCCGCGCCTGGAACGACCGACCGACCGCCTGGAAACAGCG	286
Db	1819	CGCCGCGCGGTGCAACATCTACGTGAGGCGCCTCGTGGCCCGCGGCGCGCGCGCTTCT	1878

[illegible]

**RESULT 15**

```

US-11-075-185-36
/ Sequence 36, Application US/11075185
/ Publication No. US2005026643A1
/ GENERAL INFORMATION:
/ APPLICANT: REEVES, CHRISTOPHER D
/ APPLICANT: JULIEN, BRYAN
/ APPLICANT: REID, RALPH
/ TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
/ FILE REFERENCE: 010099.03
/ CURRENT APPLICATION NUMBER: US/11/075,185
/ CURRENT FILING DATE: 2005-03-07
/ PRIOR APPLICATION NUMBER: US 60/551,103
/ PRIOR FILING DATE: 2004-03-08
/ PRIOR APPLICATION NUMBER: US 60/568,290
/ PRIOR FILING DATE: 2004-05-04
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 36
/ LENGTH: 5679
/ TYPE: DNA
/ ORGANISM: Sorangium cellulosum
US-11-075-185-36

```

Query Match	9.7%	Score 58.2;	DB 14.	Length 5679;
Best Local Similarity	46.4%;	Pred. No. 0.00042;		
Matches 189;	Conservative 0;	Mismatches 218;	Indels 0;	Gaps 0;

172 CCTGGCAACCCCGTCGACCTCGGGCGGTCCTCCCTGGCCGGACGGTCTGTGTGACCG 231

Db 3 776 CCGAGGAGCAGTCTGCTCCGCGGTGACCGTGCCTCGCGGCGCGCTGGGCGTGCAGCG 3 835

232 GAGGCGCAGTACCCGCGGCCTGGACCGGACGTCCTGGAACAGCGCGGGG 291

Age group	Number of people
0-4	180
5-9	170
10-14	160
15-19	150
20-24	140
25-29	130
30-34	120
35-39	110
40-44	100
45-49	90
50-54	80
55-59	70
60-64	80
65-69	60
70-74	50
75-79	40
80-84	30
85-89	20
90-94	10
95-99	5

.....

A vertical ruler with markings from 0 to 2.2 cm. The 0.2 cm mark is labeled '2.2 mm'.

Db 3896 GGCCTGGTCTGACCTGACGGCGGGCGAATGCAAGCGCGGTCTCTCGCGCGCGATGCAAGCG 3953

352 TCGACGGCACCGCCCTGTCCACTGTGTCTCTGCTCGCGCTCGCCGAGGCGGTCTG 411

Db 3956 CGGTGCGCGCGCGCTGGCCCTGGTGCAGGGGTGGCTCGCCGCGCCGCAAGCTGGCGGCGA 4015

412 TCGACGACCCCGAGCCTGGACACCCCTCGCGTTGGTCCAGGCGCTCGGCGCAGCCGGATCG 471

[illegible]

.....

100

[illegible]

QY 532 ATCCGCCAGGCCATGATCGATGAGCTCGAGCCGAGTGTGGCGGTG 578  
Db 4136 GCGGGTCCGGGTCTGAGCTGGGGCCGAGCCGCTGAGACGGGGCG 4182

Search completed: April 7, 2006, 03:50:12  
Job time : 228.15 secs

is Page Blank (uspto)







	Query March	20.74%	Score 124;	DB 5;	Length 60196;
	Best Local Similarity	55.04%	Pred. No. 1.e-21;		
	Matches 350;	Conservative 0;	Mismatches 250;	Indels 36;	Gaps 4;
QY	1	CCGCGCTCGCCGAGCGTACACCCGAGGAGGTGAGGTGCACT-----GGCGTACCGCAG	54		
Db	38933	CTTCGAGCGCGCGAGGCGCTTACCCAGAGGAGGTGCGCGTGGACTGTCCCGGCGCTGTGGCCG	38929		
QY	55	TGGGTGAGGGACGCGCCGCTGACCTGTCCGCTTACACCGTTCACACACAGAACTTCTGCG	114		
Db	38993	GGGCGCGCGCGCTGTGCGCGTGGACTGCGCCACGTACCGCTTTCAGCGGCAACGTACTGCG	39052		
QY	115	TCCCGGTCGCCCTTGAGGCGGAGGTCCCCGACA---CGGCGACAGATGGCGCTTACCAACTGG	171		
Db	39053	TGAGACCCCGCGCGCGACAGGCGGAACTCCGCGCCCGGCGAGGACGCGTGGCGCTTACCCGGTGG	39112		
QY	172	CTTGGCAACCCCGTTCGACTTGGGCGGCTCTCCCTGGCGCGAGCGAGTCTTGTGTGTACCG	231		
Db	39113	GCTTGGCGGCGGCGCTTCAAGCGCACCGGCGCGCGCGCGCGACCGGTGGCTGTGTGTACCG	39172		
QY	232	GAGCGCGACGTACCCCGCGCGCTTGGACGGAAGTGTCCGCGAGCGGCGCTTGGAAACAGCGCGGAG	291		
Db	39173	GCCCGAGACAGCGCGCGGAGACTGTGTGAGGCGGCTGGCGACGCGCTTACCGCGCGGAGCG	39232		
QY	292	CGACCGCTGTGTGTGTGACACCGCGCACTCG-----CGCGGCCGAGTTGGGCGCG	339		
Db	39233	CCGAGGTCGCGCTGTGTGACCGTTCGAGCGCACGACACCGACCGGCGCGGTGTGCGCGCGT	39292		

Query Match	17.2%;	Score 103.2;	DB 7;	Length 125401;
Best Local Similarity	53.5%;	Pred. No. 1.3e-16;		
Matches 291;	Conservative 0;	Mismatches 238;	Indels 15;	Gaps 3;
QY	72	GTGACCTGCGCGGTACCCGTTTCACACGACGAACTTCTGCTCCCGGTCCCTCGGCG	131	
Db	43840	GCGGACCAAGTCGGCGCCCGGTGCTGCGGACATGTGCGCTGGCGGACCCGGCGCGGACCC	43839	
QY	132	CGGGTCCCCGACACCGGCGACGATGAGCGTTACCACTCGGCTGGCACCCCTCGACCTTC	191	
Db	43900	ACATCGGCGACCAACACCCCTCGCGGACCGGGAGAGTTGGAAACCGGTGCTGCCCGCG	43959	
QY	192	GGGCGGTCCTCCCTGCGCGGACGGGTCCTGTGAGTGAACGGAGCG---GCAATACCCCCG	248	
Db	43860	AACGCCGACACACCGGCGCGGTCTCTGTGTCTGTGTCGCCGCGCGGACACACGCCCTGG	44019	
QY	249	GCTCGACGGAAGCTGTGTCGCGGACGGGCTTGAAACAGCGCGGGGCAACCTGTGTTGTTC	308	

Db 44020 GTGCGACGATGCTGCGCCGCTGCGCCGACCGCCGCGGTGAGAGTCCCGGACGAC 44079  
Qy 309 ACCGCGGACGTCGCGCGCGCGCGGATGCGGCGCGACACTGACACCGCTGACGCGACCGCTG 368  
Db 44080 GGCACCGACCGCGCGCGCGCTGCGCGCCCTGCTGACCGAAGCGCGACGACCGCGCCG 44139  
Qy 369 TCCATGTCGTCCTCTGTCGCGCTGCGCGCGCGCGGTGTCGACGA-----C 419  
Db 44140 ACCGCGGTGTCCTCCCTGCTGCGCTGACGAGACGAGCGCGCGACGAGCGGTACCGGCG 44199  
Qy 420 CCCAGCTGACACCTTCGCGTTCGAGCGCGCTCGCGCGACCGCGGATGACGTCCG 479  
Db 44200 GGCACACCGCGCACCGCGCGCTGTCAGGCGCTCGCGCGACCGCGCGCGCGCGCG 44259  
Qy 480 CTGTCGTCGTCGACGACG 536  
Db 44260 CTGTCGCGCGCTGACCG 44319  
Qy 537 GCGCGGACGATGTCGTCG 596  
Db 44320 GCGCGGCG 44379  
Qy 597 GGTG 600  
Db 44380 GCGG 44383

## RESULT 6

US-10-205-032-15  
; Sequence 15, Application US/10205032  
; Publication No. US20030113874A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Yang, Xianhu  
; APPLICANT: Stafla, Alfredo  
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN  
; FILE REFERENCE: 3016-205  
; CURRENT APPLICATION NUMBER: US/10/205,032  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 11238  
; TYPE: DNA  
; ORGANISM: micromonospora carbonacea subspecies aurantiaca  
US-10-205-032-15

Query Match 16.7%; Score 100.2; DB 5; Length 11238;

Best Local Similarity 52.9%; Pred. No. 1.2e-15;  
Matches 250; Conservative 0; Mismatches 208; Indels 15; Gaps 1;

Qy 143 CACCGGCGACGATGCGCTTACGAGCTGCGCTGCGACCCCGTGCACCTTCGCGCGGTCTC 202  
Db 2949 CACCTGAAACCGTTGCGGACGACGACCCCGACCTTCAACACCGCTGCTGAT 3008  
Qy 203 CCGCGCGGAGCGGCTGCTGTCGTCGACGCGGCGGACGACCGCGCGCTGAGAGAGT 262  
Db 3009 GGTGCGGAGCGGCTGCGCGGTCGACGCGGTCGCGGCGGACGCGCGCGCTGAGAGAGT 3068  
Qy 263 GGTGCGGAGCGGCTGCGGACGAGCGGCGGCGACCGTCTGTCGTCGACGCGCGAGTTCG 322  
Db 3069 GGTGCGGAGCGGCTGCGGACGAGCGGCGGCGACCGTCTGTCGTCGACGCGCGAGTTCG 3128  
Qy 323 GCGCGGATGCGCGCGCGCTGTCGACGCGCTGACGCGACCGCGCTGTCGTCGTCGTC 382  
Db 3129 GCTCTGACGTCG 3188  
Qy 383 TCTGTCGTCGTCG 428  
Db 3189 CTTGTCGTCGTCG 3248  
Qy 429 -GACACCGCTTCGCGGTCGACGCGCTCGCGCGACCGCGGATGACGTCCTCCCTGTCG 487

Db 3249 CGCTGCTTCGCTTACCTTACGAGCATTTGCGGACGCGGAGACGACGCGCGCTTCGCG 3308  
Qy 488 GGTGACGAGGAGCG 547  
Db 3309 GGTGACGAGGAGGCG 3368  
Qy 548 GGTGCGGTCG 600  
Db 3369 GGTGTCGCGGCTTCG 3421

## RESULT 7

US-10-819-386A-1/c  
; Sequence 1, Application US/10819386A  
; Publication No. US2005089884A1  
; GENERAL INFORMATION:  
; APPLICANT: Korea Advanced Institute of Science and Technology  
; APPLICANT: Shanghai Jiaotong University  
; APPLICANT: LEE, Sang Yup  
; APPLICANT: DENG, ZIXIN  
; APPLICANT: CHEN, SHI  
; APPLICANT: JERONG, Xi Jun  
; APPLICANT: ZHOU, XIUPEN  
; TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polyketides  
; FILE REFERENCE: P003-B015  
; CURRENT APPLICATION NUMBER: US/10/819,386A  
; CURRENT FILING DATE: 2004-04-06  
; PRIOR APPLICATION NUMBER: KR10-2003-0074035  
; PRIOR FILING DATE: 2003-10-23  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 138203  
; TYPE: DNA  
; ORGANISM: Streptomyces sp. FR-008  
US-10-819-386A-1

Query Match 16.7%; Score 100.2; DB 9; Length 138203;  
Best Local Similarity 53.7%; Pred. No. 6.9e-16;  
Matches 341; Conservative 0; Mismatches 258; Indels 36; Gaps 5;

Qy 2 CGCGCTGCGCGGCGTACACCGCGGCGGTGAGAGTGCATGCGGCTACCGGATGAGTGA 61  
Db 92504 CCGCTGACGCGCGCTTTCGACCGCGCTGACGCGGAGGACCTGACCGCTTCGCGAC 92445  
Qy 62 GGAAGCGCGCGTTCACCGCGCGGTTCACCGGTCGACGAGAACCTTCGCGCGGT 121  
Db 92444 GACCTGCGGTGACACCGAGCGCTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCG 92385  
Qy 122 CCGCTGCGCGGTTCGCGCGGACCGAGCGGACCGAGCGGTCGCGCGGTTCGCGCGGT 179  
Db 92384 GCGCGCGCGCGGACCGAGCGGACCGAGCGGTCGCGCGGTTCGCGCGGTTCGCG 92325  
Qy 180 CCGTTCGACCTGCGCGGTTCCTCTGCGCGGTTCGCGCGGTTCGCGCGGTTCGCG 232  
Db 92324 GCTGTCGTCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 92265  
Qy 233 AGCGGAGTACCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 292  
Db 92264 GCGGTTCGCGCGGACCGCTGAGACCGCGCGGTTCGCGCGGTTCGCGCGGTTCG 92205  
Qy 293 GACGTCGTCGTCGACCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCG 349  
Db 92204 CCGCTTCACCGTTCGACCGCTTCGAGCGGACCGCGGTTCGCGCGGTTCGCGCGGT 92145  
Qy 350 CGTTCGCGCGACCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 400  
Db 92144 CGCGGAGCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 92085  
Qy 401 -----GCGCGGTTCGTCGAGAACCGCGGTCGAGAACCGCTTCGCGGTTCG 448  
Db 92084 GCGAGCGCTTCGCGCGACCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 92025

QY 449 GGCGCTGGGGCGAAGCCGGGATTCACACGTCCTCTGGTGTATACCAAGGAGACGGCCGCC 508

Db 92024 GGCCCTCTCGAGCGCCCGGGGTTCACGCGCCCTGTGGTGTCTTACACCGGGAGACCGTCTGC 91965

QY 509 CGTACACCGT---CGGAGACGACGTCGATTCGGGCGCAAGGCCATGTCGTGGGCTCGGCCG 565

Db 91964 GGTGGCGGGGTTCCGAGCGCGTCACCGGCCCGCCAGGCGCGCGTCTGGGGCGCTGGGCCG 91905

QY 566 GGATGGAGCGTGGAGTCCCGCGCCCGCGTGGGGTG 600

Db 91904 GGTGCGCGCGCTCGAACAACCCCGCGCTGTGGGGCG 91870

## RESULT 8

```

US-10-042665A-3
; Sequence 3, Application US/10042665A
; Publication No. US20050053927A1
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Toupet, Christine
; APPLICANT: Engel, Nathalie
; TITLE OF INVENTION: Rifamycin biosynthesis gene cluster
; FILE REFERENCE: 4-21001/E/C1
; CURRENT APPLICATION NUMBER: US/10/042,665A
; CURRENT FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: 09/242,744
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: PCT/EP97/04495
; PRIOR FILING DATE: 1997-08-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 53799
; TYPE: DNA
; ORGANISM: Amycolatopsis mediterranei
; US-10-042-665A-3

```

[illegible]

QY	599	TG	600
Db	14079	CG	14080

## RESULT 9

```

US-09-861-289-30
Sequence 30, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 13842
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-30

```

	Query Match	14.9%	Score 89.4	DB 3	Length 13842
	Best Local Similarity	52.6%	Pred. No. 5e-13		
	Matches 317	Conservative 0	Mismatches 271	Indels 15	Gaps 5
Qy	1	CCGCGCTCGCCGAGAGCTTACACCCGGGGGGGTGAGGTGCACCTGACGCAATG---G	57		
Db	5849	CTTCGCTCGCCGAGAGCGTGGGTCAACGGGGCTTCCCTGGCATGTACTTCGCTCGCCG	5908		
Qy	58	GTCAGGAGACCCCGGTTGACCTTCGCGGTCTAACCCGTTCCAACACAGACAACTTTCGCTTC	117		
Db	5909	CCAGGCGTTCGCGCCCGCGTCTGCGCCACTAACGCTTTCAGGCCGACCGCTCACTACGGCTCG	5968		
Qy	118	CGGTCGCCCTTGAGGCGCGGGTCCCGACACACGAGCAGAGTGGAGCTTTACAGCTCGCTGAC	177		
Db	5969	AGAAACATCCCCGC---GCCCTGACCACCGGACAGACTGGCGCTTACCGATCACTGAC	6025B		
Qy	178	ACCCCGTGCACCTTCGAGCGGTCTCTCTGACCGGACGAGTCTTGCTGTGACCCGAGACGG	237		
Db	6026	AGGCGCTCCGAGCGCCGAGAGGGGTCCGAGGCAACCG---CTGTCCGGCGCTGGGCTCG	6082		
Qy	238	CAGTACCCCGCGGCTGTGACGGAAGTGTCCGCGACCGGCTTGAGAACAGCGCGGGCGACCG	297		
Db	6083	CCGTCAACGCGGAGAGACACTTCGCGGAGGCGCGCGCGTGCCTCAACGCGCTGTGTGACG	6142		
Qy	298	TGCTGTGTGACACCGGCGAGTCCGCGCCCGGATTCGAGCGCGCACTGACGCGCTGCAGC	357		
Db	6143	CCGGGCGGAAAGTGTGAGGTGTCTACCGCCGGGGCGGACAGACGACCTGAGGCTTCGCG	6202		
Qy	358	GCAACCGCCGTGTCACTGTGTCTCTCTGTCTGCGCGCTCGCCGAGGCGGTGTGTGACG	417		
Db	6203	---CCGGGCTCAACCGCACTGACGACCGGTGACGAGCTTCAACCGCGTGTGTCTCGCTCTCG	6259		
Qy	418	ACCCGAGCGCTGACACACTCTCGCGTGTGTCCAGGCGCTCGGCGGACGCCGGATGTGACGTC	477		
Db	6260	ACGGAATCTGTACCGCAGGTGTGCTTGGGTCCAGGCGGCTCGGAGACGCCGGAATCAAGCGCG	6319		
Qy	478	CCCTGTGGCTGTGACACGAGGACGCGCGCGCTGTACCG---TCGAGACGACGTGCATC	534		
Db	6320	CCCTGTGTGTCCGTCAACCGAGGGCGGCGTCTCGTCCGAGGTCTGCACACCCCGCGGACG	6379		
Qy	535	CGGCGCCAGAGCTATGTCGTGTGGGCTTCGCGGGGTGTGGCGTGTGAGTCCCCCGCCGCT	594		
Db	6380	CCGACCGGGCGCAAGTCTTGAGGCGCTTCGCGCGCGTGTGTGCGCTTTGAGCACCCCGAACGCT	6439		

QY 595 GGG 597  
Db 6440 GGG 6442

RESULT 10  
US-09-860-846-30  
; Sequence 30, Application US/09860846  
; Patent No. US20020164742A1  
GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/860,846  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 13842  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-30

Query Match 14.9%; Score 89.4; DB 3; Length 13842;  
Best Local Similarity 52.6%; Pred. No. 5e-13;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTGCGGAGGCGTACACCGGGGGGTGAGGTGACGTGACCGCAGTG---G 57  
Db 5849 CCTGCTGCGGAGGCGTGTGACCGGGCTTCCGTGGCACTGCTGCTGCGCCG 5908  
QY 58 GTGAGGAGACGCGCGGTGACCTGCGGTCTACCGGTTCAACGACAGAACTTTGAGCTCC 117  
Db 5909 CCAGGCGCTCCGCGCGGTGCTGCGCACTTACGCTTCAAGGCGAGGCTACTAGGCTCG 5968  
QY 118 CGGTCCCTGAGGCGGGGTCCCGGACACCGGCGACGAGTGGCGTTACCAAGCTCGCTGGC 177  
Db 5969 AGAACACTCCCGC---GCCCTGGGCACCGGCGACGACTGGCGGCTACCGCACTGCACTGGA 6025  
QY 178 ACCCGGTGACCTGAGGCGGTCTCCCTGGCGGACCGGCTGCTGAGTGGTGAACCGGAGCG 237  
Db 6026 AGCGCTCCGCGCGCGGAGGAGGTCCAGCGCACCGG---CCTGTCCGAGCGTGGCTCG 6082  
QY 238 CAGTACCCCGGCGCTGACGAGACGTGTCCGCGACGCGCTGGAACAGCGCGGAGCGACCG 297  
Db 6083 CGGTACGCGCGGAGGACCACTCCGCGGACAGGCGCGCGCTGCTCACCGCGCTGGTCAGCG 6142  
QY 298 TCGTGTGTGACACCGCGACGTGCGCGCCCGGATCGGCGCGGCACTGACCGCGTGAAG 357  
Db 6143 CCGGAGGAGAGTGAAGTGTGACGCGCGGAGGAGGAGACGACGACCTGTGAGGCGCTTCGCG 6202  
QY 358 GCAACGCGCTGTGCACTGTGTGCTCTGCTGCGGCTGCGGAGGCGGAGTGTGTCAGAG 417  
Db 6203 ---CCGGCTACCGCACTGACGACCGGTGACGCGCTTACCGGCGTGTCTGCTGCTCG 6259  
QY 418 ACCCAGCTGAGACACCTCGCGTGTGTCAGGCGCTCGGCGACGCGGAGATGACGCTCC 477  
Db 6260 ACGACTGTACCGGAGGTGCGCTGGGTTCAGGCGCTCGGCGACCGCGGATCAAGGCGCG 6319  
QY 478 CCTGTGCTGTGTAACAAGGACGCGCGCGCGTGAACG---TGGAGACGACGTGATC 534  
Db 6320 CCTGTGCTGTGTAACAAGGAGGCGGAGTCTCGTGGACGACGTCTGACACACCGCGCGACG 6379  
QY 535 CGGCGCAGGCGCATGTGTGCTGAGGCTCGGCGGAGTGTGAGGCTGAGAGTCCCGCGCGGCT 594  
Db 6380 CGGACCGGCGCATGTCTGTGGGCGCTCGGCGGCTGTGCGCTTGAAGACCCCGAAGCGCT 6439  
QY 595 GGG 597

Db 6440 GGG 6442

RESULT 11  
US-09-988-384B-30  
; Sequence 30, Application US/0988384B  
; Publication No. US20030073824A1  
GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.536US1  
; CURRENT APPLICATION NUMBER: US/09/988,384B  
; PRIOR FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: PCT/US99/14398  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,537  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 30  
; LENGTH: 13842  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-988-384B-30

Query Match 14.9%; Score 89.4; DB 3; Length 13842;  
Best Local Similarity 52.6%; Pred. No. 5e-13;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTGCGGAGGCGTACACCGGGGGGTGAGGTGACGTGACCGCAGTG---G 57  
Db 5849 CCTGCTGCGGAGGCGTGTGACCGGGCTTCCGTGGCACTGCTGCTGCGCCG 5908  
QY 58 GTGAGGAGACGCGCGGTGACCTGCGGTCTACCGGTTCAACGACAGAACTTTGAGCTCC 117  
Db 5909 CCAGGCGCTCCGCGCGGTGCTGCGCACTTACGCTTCAAGGCGAGGCGCTACTAGGCTCG 5968  
QY 118 CGGTCCCTGAGGCGGGGTCCCGGACACCGGCGACGAGTGGCGTTACCAAGCTCGCTGGC 177  
Db 5969 AGAACACTCCCGC---GCCCTGGGCACCGGCGACGACTGGCGGCTACCGCACTGCACTGGA 6025  
QY 178 ACCCGGTGACCTGAGGCGGTCTCCCTGGCGGACCGGCTGCTGAGTGGTGAACCGGAGCG 237  
Db 6026 AGCGCTCCGCGCGCGGAGGAGGTCCAGCGCACCGG---CCTGTCCGAGCGTGGCTCG 6082  
QY 238 CAGTACCCCGGCGCTGACGAGACGTGTCCGCGACGCGCTGGAACAGCGCGGAGCGACCG 297  
Db 6083 CGGTACGCGCGGAGGACCACTCCGCGGACAGGCGCGCGCTGCTCACCGCGCTGGTCAGCG 6142  
QY 298 TCGTGTGTGACACCGCGACGTGCGCGCCCGGATCGGCGCGGCACTGACCGCGTGAAG 357  
Db 6143 CCGGAGGAGAGTGAAGTGTGACGCGCGGAGGAGGAGACGACGACCTGTGAGGCGCTTCGCG 6202  
QY 358 GCAACGCGCTGTGCACTGTGTGCTCTGCTGCGGCTGCGGAGGCGGAGTGTGTCAGAG 417  
Db 6203 ---CCGGCTACCGCACTGACGACCGGTGACGCGCTTACCGGCGTGTCTGCTGCTCG 6259  
QY 418 ACCCAGCTGAGACACCTCGCGTGTGTCAGGCGCTCGGCGACGCGGAGATGACGCTCC 477  
Db 6260 ACGACTGTACCGGAGGTGCGCTGGGTTCAGGCGCTCGGCGACCGCGGATCAAGGCGCG 6319  
QY 478 CCTGTGCTGTGTAACAAGGACGCGCGCGCGTGAACG---TGGAGACGACGTGATC 534  
Db 6320 CCTGTGCTGTGTAACAAGGAGGCGGAGTCTCGTGGACGACGTCTGACACACCGCGCGACG 6379  
QY 535 CGGCGCAGGCGCATGTGTGCTGAGGCTCGGCGGAGTGTGAGGCTGAGAGTCCCGCGCGGCT 594  
Db 6380 CGGACCGGCGCATGTCTGTGGGCGCTCGGCGGCTGTGCGCTTGAAGACCCCGAAGCGCT 6439  
QY 595 GGG 597

Db 6440 GGG 6442

```
RESULT 12
US-09-836-821-30
/ Sequence 30, Application US/09836821
/ Publication No. US20030087405A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA encoding methymycin and pikromycin
/ FILE REFERENCE: 600.438US1
/ CURRENT APPLICATION NUMBER: US/09/836,821
/ PRIOR FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: US 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 43
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 30
/ LENGTH: 13842
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-836-821-30
```

Query Match 14.9%; Score 89.4; DB 3; Length 13842;  
Best Local Similarity 52.6%; Pred. No. 5e-13;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

```
QY 1 CCGCGCTCCGCGGAGGTACACCGGGGGGTGAGGTGACCTGCGCTACCGCAGTG---G 57
Db 5849 CTTCCGCTCCGCGGAGGTGAGGTGACCGGGCTTCCCGGAGTACCTTCCCTCCGCGG 5908
QY 58 GTGAGGAGCGCCCGGTGACCTGCGGCTTACCCGTTCAACGACAGAACTTGTGGCTCC 117
Db 5909 CCAAGGCTCCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 5968
QY 118 CGGTCCCGTCCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 177
Db 5969 AGAACAATCCCGGCG---GCCCTGGGCAACCGGCGACGACCTGCGCTACCGCATGACCTGA 6025
QY 178 ACCCGGTGACCTCCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 237
Db 6026 AGCGCTCCCGGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6082
QY 238 CAGTACCCCGGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 297
Db 6083 CGGTACCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6142
QY 298 TCGTGTGTGACCGGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 357
Db 6143 CCGGGGCGAAGGTGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6202
QY 358 GCAAGCGGCTTCCCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 417
Db 6203 ---CCGCGCTCAACCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6259
QY 418 ACCCGAGCTGAGCAACCTCCGCTTGTGTCAGGCGCTCCGCGGAGTACCTTCCCTCCGCGG 477
Db 6260 ACGAGCTGTCAGGAGGTGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6319
QY 478 CCTGTGTGTGACCGGCGGAGGTGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 534
Db 6320 CCTGTGTGTGACCGGCGGAGGTGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6379
QY 535 CCGGCGGAGCTACGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 594
Db 6380 CCGGCGGAGCTACGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 6439
QY 595 GGG 597
|||
```

Db 6440 GGG 6442

```
RESULT 13
US-10-271-889-30
/ Sequence 30, Application US/10271889
/ Publication No. US20030194784A1
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D.H.
/ APPLICANT: Liu, H.
/ APPLICANT: Xue, Y.
/ APPLICANT: Zhao, L.
/ TITLE OF INVENTION: DNA Encoding Methymycin and Pikromycin
/ FILE REFERENCE: 600.582US1
/ CURRENT APPLICATION NUMBER: US/10/271,889
/ PRIOR FILING DATE: 2002-10-15
/ PRIOR APPLICATION NUMBER: US 09/861,289
/ PRIOR FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: US 09/860,846
/ PRIOR FILING DATE: 2001-05-18
/ PRIOR APPLICATION NUMBER: US 09/836,821
/ PRIOR FILING DATE: 2001-04-17
/ PRIOR APPLICATION NUMBER: US 09/105,537
/ PRIOR FILING DATE: 1998-06-26
/ NUMBER OF SEQ ID NOS: 55
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 30
/ LENGTH: 13842
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-10-271-889-30
```

Query Match 14.9%; Score 89.4; DB 6; Length 13842;  
Best Local Similarity 52.6%; Pred. No. 5e-13;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

```
QY 1 CCGCGCTCCGCGGAGGTACACCGGGGGGTGAGGTGACCTGCGCTACCGCAGTG---G 57
Db 5849 CTTCCGCTCCGCGGAGGTGAGGTGACCGGGCTTCCCGGAGTACCTTCCCTCCGCGG 5908
QY 58 GTGAGGAGCGCCCGGTGACCTGCGGCTTACCCGTTCAACGACAGAACTTGTGGCTCC 117
Db 5909 CCAAGGCTCCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 5968
QY 118 CGGTCCCGTCCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 177
Db 5969 AGAACAATCCCGGCG---GCCCTGGGCAACCGGCGACGACCTGCGCTACCGCATGACCTGA 6025
QY 178 ACCCGGTGACCTCCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 237
Db 6026 AGCGCTCCCGGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6082
QY 238 CAGTACCCCGGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 297
Db 6083 CGGTACCGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6142
QY 298 TCGTGTGTGACCGGCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 357
Db 6143 CCGGGGCGAAGGTGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6202
QY 358 GCAAGCGGCTTCCCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 417
Db 6203 ---CCGCGCTCAACCGGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6259
QY 418 ACCCGAGCTGAGCAACCTCCGCTTGTGTCAGGCGCTCCGCGGAGTACCTTCCCTCCGCGG 477
Db 6260 ACGAGCTGTCAGGAGGTGAGGTGAGGTGACCGGCTTCCCGGAGTACCTTCCCTCCGCGG 6319
QY 478 CCTGTGTGTGACCGGCGGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 534
Db 6320 CCTGTGTGTGACCGGCGGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 6379
QY 535 CCGGCGGAGCTACGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 594
|||
```

Db 6380 CCGACCGGCGCATGCTCTGGGGGCTCTGGCGCGCTGCTGCGCTTGAAGACCCCGAAGCT 6439  
QY 595 GGG 597  
Db 6440 GGG 6442

## RESULT 14

US-09-861-289-5  
; Sequence 5, Application US/09861289  
; Patent No. US20020110897A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600,438US1  
; CURRENT APPLICATION NUMBER: US/09/861,289  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-861-289-5

Query Match 14.9%; Score 89.4; DB 3; Length 36778;  
Best Local Similarity 52.6%; Pred. No. 4,1e-13;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 57  
Db 7590 CCGGCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 7649  
QY 58 GTGAGGAGCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 117  
Db 7650 CCACGCGCTCCCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 7709  
QY 118 CGGTCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 177  
Db 7710 AGAAGACCTCCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 7766  
QY 178 ACCCGGTGACCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 237  
Db 7767 AGCGCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 7823  
QY 238 CAGTACCCCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 297  
Db 7824 CCGTACCGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 7883  
QY 298 TCGTGTGTGACCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 357  
Db 7884 CCGGCGGAGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 7943  
QY 358 GCAACGCGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 417  
Db 7944 ---CCCGCTACCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 8000  
QY 418 ACCCGAGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 477  
Db 8001 ACAGACTGTACCGGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 8060  
QY 478 CCGTGTGTGAGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 534  
Db 8061 CCGTGTGTGAGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 8120  
QY 535 CCGGCGGAGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 594  
Db 8121 CCGGCGGAGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 8180

Db 8121 CCGACCGGCGCATGCTCTGGGGGCTCTGGCGCGCTGCTGCGCTTGAAGACCCCGAAGCT 8180  
QY 595 GGG 597  
Db 8181 GGG 8183

## RESULT 15

US-09-860-846-5  
; Sequence 5, Application US/09860846  
; Patent No. US20020164742A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600,438US1  
; CURRENT APPLICATION NUMBER: US/09/860,846  
; CURRENT FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: 09/105,537  
; PRIOR FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5  
; LENGTH: 36778  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-860-846-5

Query Match 14.9%; Score 89.4; DB 3; Length 36778;  
Best Local Similarity 52.6%; Pred. No. 4,1e-13;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;

QY 1 CCGGCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 57  
Db 7590 CCGGCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 7649  
QY 58 GTGAGGAGCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 117  
Db 7650 CCACGCGCTCCCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 7709  
QY 118 CGGTCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 177  
Db 7710 AGAAGACCTCCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 7766  
QY 178 ACCCGGTGACCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 237  
Db 7767 AGCGCTCTGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 7823  
QY 238 CAGTACCCCGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 297  
Db 7824 CCGTACCGCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 7883  
QY 298 TCGTGTGTGACCGCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 357  
Db 7884 CCGGCGGAGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 7943  
QY 358 GCAACGCGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 417  
Db 7944 ---CCCGCTACCGGAGCTTACACCCGCGGGGTGAGGTGACTGCGCTGACCGAAGT---G 8000  
QY 418 ACCCGAGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 477  
Db 8001 ACAGACTGTACCGGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 8060  
QY 478 CCGTGTGTGAGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 534  
Db 8061 CCGTGTGTGAGCTGACCTGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 8120  
QY 535 CCGGCGGAGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 594  
Db 8121 CCGGCGGAGAGTGTGAGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACAACTTTGAGCTCC 8180

Qy	595	GGG	597
Db	8181	GGG	8183

Search completed: April 7, 2006, 04:35:24  
Job time : 424.485 secs



**inis Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 01:33:46 ; Search time 342.081 Seconds (without alignments)

11689.676 Million cell updates/sec

Title:	US-10-611-442-2_COPY_1_600
Perfect score:	600
Sequence:	1 ccgcgcctcgccgagcgctac.....gtcccccgcgcggtggggtg 600

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues  
Total number of hits satisfying chosen parameters

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%

### Listing first 45 summaries

1: N\_Geneseq.21.\*  
2: geneseqn1980s.\*  
3: geneseqn1990s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002as.\*  
7: geneseqn2002bs.\*  
8: geneseqn2003as.\*  
9: geneseqn2003bs.\*  
10: geneseqn2003cs.\*  
11: geneseqn2003cds.\*  
12: geneseqn2004as.\*  
13: geneseqn2004bs.\*  
14: geneseqn2005as.\*

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	600	100.0	17566	12	AD114148	Ad114148 M. megala
2	600	100.0	47961	4	AAE30757	AAE30757 Micromono
3	286.6	47.8	29879	2	AAQ64686	AAQ64686 eryA regi
4	189	31.5	9510	12	ADAM97245	Adm97245 Synthetic
5	124	20.7	4275	10	AAAD55818	Adad55818 Micromon
6	124	20.7	60136	10	ADAD5810	Adad5810 Micromon
7	103.2	17.2	125401	4	AAAD17186	Adad17186 Streptom
8	100.2	16.7	11238	14	AAAD55817	Adad55817 Micromon
9	100.2	16.7	138203	14	ADAD51725	Adad51725 PR-008 po
c	98	16.3	53789	2	AAAZ21187	AAAZ21187 Amycolato
10	98	14.9	13842	3	AAAZ67297	AAAZ67297 S. venezu
11	89.4	14.9	13842	12	ADL91915	ADL91915 Streptomu
12	89.4	14.9	36778	3	AAAZ87318	AAAZ87318 S. venezu
13	89.4	14.9	36778	12	ADL91913	ADL91913 Streptomu
14	89.4	14.9	37948	3	AAAZ87285	AAAZ87285 S. venezu
15	89.4	14.9	38506	3	AAAZ75633	AAAZ75633 Nucleotid
16	89.4	14.9	38506	3	AAAZ56001	AAAZ56001 Recombina
17	89.4	14.9	38506	8	ADAO99418	ADAO99418 Comaid pk
18	89.4	14.9	38506	10	ADH53462	ADH53462 S. venezu
19	89.4	14.9	38506	10	ADH53462	ADH53462 S. venezu

20	89.4	14.9	38560	10	AB556090	AB556090 S. venezue
21	87.2	14.5	2325	11	AD123088	AD123088 Streptomy
22	84.8	14.1	4689	3	AA287299	AA287299 S. venezue
23	84.8	14.1	4689	12	ADL91919	ADL91919 Streptomy
24	84	14.0	11220	3	AA287298	AA287298 S. venezue
25	84	14.0	11220	12	ADL91917	ADL91917 Streptomy
26	82	13.7	113193	8	ADP54645	ADP54645 Streptomy
27	81	13.5	67167	14	ADY80053	ADY80053 Nucleotid
28	80.4	13.4	74787	12	ADQ74672	ADQ74672 Streptomy
29	79.6	13.3	77536	3	AA14651	AA14651 Nucleotid
30	79.2	13.2	4770	8	AA161172	AA161172 Actinosyn
31	79.2	13.2	82746	8	AA161122	AA161122 Actinosyn
32	77.6	12.9	27541	4	AD171785	AD171785 Streptomy
33	74.8	12.5	65140	4	AD17184	AD17184 Streptomy
34	74.8	12.5	73984	14	ADY26824	ADY26824 S. pulvere
35	74.6	12.4	50937	3	AAA09469	AAA09469 Streptoco
36	74.4	12.4	9684	13	ADQ91711	ADQ91711 Polyketid
37	74.4	12.4	9684	13	AE87003	AE87003 Streptomy
38	74.4	12.4	9684	13	AE86800	AE86800 Streptomy
39	74.4	12.4	164051	13	ADQ91695	ADQ91695 Polyketid
40	74.4	12.4	164051	13	AE86862	AE86862 Streptomy
41	74.4	12.4	164051	13	AE86652	AE86652 Streptomy
42	73.8	12.3	9795	14	ADY72595	ADY72595 A. orient
43	73.8	12.3	73599	14	ADY72575	ADY72575 Polyene p
44	72.2	12.0	7788	10	AD123894	AD123894 Streptomy
45	72.2	12.0	37360	10	AD123892	AD123892 Streptomy

## ALIGNMENTS

	RESULT 1
ID	AD114148
XX	AD114148 standard; DNA; 17596 BP.
AC	AD114148;
DT	15-APR-2004 (first entry)
DE	M. megalomicea cosmid KOS205-57-2.3B SEQ ID NO:2.
KW	ds; polypeptide; enzyme; MegR; MegI; MegK; MegCV; MegBYI; MegBIT;
XX	MegU; MegM; megosamine; megalomicin.
XX	Micromonospora megalomicea.
PN	WO2004003169-A2.
PD	08-JAN-2004.
PP	30-JUN-2003; 2003WO-US020681.
PR	28-JUN-2002; 2002US-0393016P.
PA	(KOSA-) KOSAN BIOSCIENCES INC.
FI	Hutchinson RC, Katz L, Reid R, Hu Z, Gramajo H,
DR	WPI; 2004-203379/19.
PT	Novel isolated, purified, or recombinant nucleic acid comprising
PT	polypeptide modifying gene, where gene encodes polypeptide modifying enzyme
PT	e.g., MegR, MegK, or MegM enzymes useful for producing modified
PS	polyketide.
XX	Example 2; SEQ ID NO 2; sipp; English.
CC	The invention relates to a novel isolated, purified, or recombinant
CC	nucleic acid (I) comprising a polypeptide modifying gene, where the gene
CC	encodes a polypeptide modifying enzyme chosen from MegR, MegI, MegK,
CC	MegCV, MegCU, MegBYI, MegBIT, MegL, and MegM enzymes. A method of the
CC	invention is useful for producing a modified polypeptide, which involves
CC	culturing a recombinant cell comprising the recombinant nucleic acid

under conditions in which the cell expresses a product of a gene encoded by the nucleic acid under conditions in which the unmodified polyketide is present, and producing the modified polyketide. The cell produces megalomycin, and can attach megalomycin to a polyketide, where the cell, it is naturally occurring non-recombinant state cannot produce megalomycin. The present sequence contains downstream megalomycin modification enzyme genes.

Sequence 17596 BP; 2325 A; 5966 C; 6604 G; 2701 T; 0 U; 0 Other;

Query Match 100.0%; Score 600; DB 12; Length 17596;

Best Local Similarity 100.0%; Pred. No. 4.3e-85;

Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCGGCGCTCCGCAAGCGTCAACCCGGGGGTGGAGGTGCACTGGCGTACCGCAATGGGTG 60
    1 CCGGCGCTCCGCAAGCGTCAACCCGGGGGTGGAGGTGCACTGGCGTACCGCAATGGGTG 60
DB 1 CCGGCGCTCCGCAAGCGTCAACCCGGGGGTGGAGGTGCACTGGCGTACCGCAATGGGTG 60
QY 61 AGGAGCGCCGGGTGCACTGCGGGTCTACCGGTTCAACGACAGAACTTCTGGCTCCGG 120
    61 AGGAGCGCCGGGTGCACTGCGGGTCTACCGGTTCAACGACAGAACTTCTGGCTCCGG 120
DB 61 AGGAGCGCCGGGTGCACTGCGGGTCTACCGGTTCAACGACAGAACTTCTGGCTCCGG 120
QY 121 TCCCGCTGGGCGGGGTCCCGCAACCGGCGACAGAGTGGCGTTACCACTGCGTGGAC 180
    121 TCCCGCTGGGCGGGGTCCCGCAACCGGCGACAGAGTGGCGTTACCACTGCGTGGAC 180
DB 121 TCCCGCTGGGCGGGGTCCCGCAACCGGCGACAGAGTGGCGTTACCACTGCGTGGAC 180
QY 181 CCGTTCGACTTGGGCGGTCTCTCTGGTCCCGGACGGGTCTCTGGTGTGAACCGGAC 240
    181 CCGTTCGACTTGGGCGGTCTCTCTGGTCCCGGACGGGTCTCTGGTGTGAACCGGAC 240
DB 181 CCGTTCGACTTGGGCGGTCTCTCTGGTCCCGGACGGGTCTCTGGTGTGAACCGGAC 240
QY 241 TACCCCGGCGCTGACGAGCGTGGTTCGCGACGCGGCTGGAACAGCGCGGCGACCGT 300
    241 TACCCCGGCGCTGACGAGCGTGGTTCGCGACGCGGCTGGAACAGCGCGGCGACCGT 300
DB 241 TACCCCGGCGCTGACGAGCGTGGTTCGCGACGCGGCTGGAACAGCGCGGCGACCGT 300
QY 301 TGTGTGACACCGCGAGTGCAGCGCCCGGATCGGCGCGCACTGACGCGCTGACGCG 360
    301 TGTGTGACACCGCGAGTGCAGCGCCCGGATCGGCGCGCACTGACGCGCTGACGCG 360
DB 301 TGTGTGACACCGCGAGTGCAGCGCCCGGATCGGCGCGCACTGACGCGCTGACGCG 360
QY 361 CCGGCGCTGACGAGTGTCTCTCTGGTCCCGGACGGGTCTCTGGTGTGAACCGGAC 420
    361 CCGGCGCTGACGAGTGTCTCTCTGGTCCCGGACGGGTCTCTGGTGTGAACCGGAC 420
DB 361 CCGGCGCTGACGAGTGTCTCTCTGGTCCCGGACGGGTCTCTGGTGTGAACCGGAC 420
QY 421 CCAGCGTGAACACCTCGCGTGTGTCAGGCGCTCGGCGCGACCGGATCGACGTCGCC 480
    421 CCAGCGTGAACACCTCGCGTGTGTCAGGCGCTCGGCGCGACCGGATCGACGTCGCC 480
DB 421 CCAGCGTGAACACCTCGCGTGTGTCAGGCGCTCGGCGCGACCGGATCGACGTCGCC 480
QY 481 TGTGCTGTGTGACCAAGGACCGCGCGCGTGAACCGTGAACGACGTCGTCGAGCC 540
    481 TGTGCTGTGTGACCAAGGACCGCGCGCGTGAACCGTGAACGACGTCGTCGAGCC 540
DB 481 TGTGCTGTGTGACCAAGGACCGCGCGCGTGAACCGTGAACGACGTCGTCGAGCC 540
QY 541 AGGCGATGTGTGTGGGTTCGCGCGGGGTGGTGGGTGGAATCCCCCGGTGGGGTG 600
    541 AGGCGATGTGTGTGGGTTCGCGCGGGGTGGTGGGTGGAATCCCCCGGTGGGGTG 600
DB 541 AGGCGATGTGTGTGGGTTCGCGCGGGGTGGTGGGTGGAATCCCCCGGTGGGGTG 600
```

RESULT 2

AAE30757 AAF30757 standard; DNA; 47981 BP.

AAE30757;

21-JUN-2001 (first entry)

Micromonospora megalomicea megalomycin biosynthetic gene cluster.

Megalomycin; meg gene; polyketide synthase; antibiotic; motilide;

antiparasitic; de.

Micromonospora megalomicea subsp. nigra.

Key Location/Qualifiers

FT CDS complement(1..144)

FT /\*tag= a

/partial  
/gene="megBVI(megT)" /product="TDP-4-keto-6-deoxyglucose-"  
/note="encodes AAB82201"  
928..2061

/\*tag= b  
/gene="megDVI"  
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
/note="encodes AAB82202"  
2072..3382

/\*tag= c  
/gene="megDI"  
/product="TDP-megosamine glycosyltransferase"  
/note="eryCII homologue; encodes AAB802203"  
3462..4634

/\*tag= d  
/gene="megI"  
/product="mycarose O-acyltransferase"  
/note="encodes AAB82204"  
4651..5775

/\*tag= e  
/gene="megDIII"  
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"  
/note="eryCI, DmrV homologue; encodes AAB82205"  
5822..6595

/\*tag= f  
/gene="megDIIII"  
/product="daunosaminyl-N-N-dimethyltransferase"  
/note="eryCVI homologue; encodes AAB82206"  
6592..7197

/\*tag= g  
/gene="megDIV"  
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"  
/note="eryBVIII, dnmV homologue; encodes AAB82207"  
7220..8206

/\*tag= h  
/gene="megDV"  
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/note="eryVII, dnmV homologue; encodes AAB82208"  
complement(8228..9220)

/\*tag= i  
/gene="megBII-1(megDVII)"  
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"  
/note="encodes AAB82209"  
complement(9226..10479)

/\*tag= j  
/gene="megBV"  
/product="TDP-mycarose glycosyltransferase"  
/note="encodes AAB82210"  
complement(10483..11424)

/\*tag= k  
/gene="megBIV"  
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/note="encodes AAB82211"  
12181..22821

/\*tag= l  
/gene="megA"  
/product="megalomycin 6-deoxyerthronolide B synthase 1"  
/note="polyketide synthase; encodes AAB82212"  
12505..13470

/\*tag= m  
/gene="megA"  
/function="AT-L"  
13576..13791

/\*tag= n  
/gene="megA"  
/function="ACP-L"  
13849..15126

/\*tag= o  
/gene="megA"  
/function="KS1"  
15427..16476

/\*tag= p

```
PT /gene= "mega"
PT /function= "AT1"
PT misc_feature
PT /tag= q
PT /gene= "mega"
PT /function= "KR1"
PT misc_feature
PT /tag= x
PT /gene= "mega"
PT /function= "ACPl"
PT misc_feature
PT /tag= b
PT /gene= "mega"
PT /function= "KS2"
PT misc_feature
PT /tag= c
PT /gene= "mega"
PT /function= "AT2"
PT misc_feature
PT /tag= u
PT /gene= "mega"
PT /function= "KR2"
PT misc_feature
PT /tag= v
PT /gene= "mega"
PT /function= "ACP2"
PT CDS
PT /tag= w
PT /gene= "megaliI"
PT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
PT /note= "polyketide synthase, encodes AAB8213"
PT misc_feature
PT /tag= x
PT /gene= "megaliI"
PT /function= "KS3"
PT misc_feature
PT /tag= ai
PT /gene= "megaliII"
PT /function= "KS5"
PT misc_feature
PT /tag= y
PT /gene= "megaliI"
PT /function= "AT3"
PT misc_feature
PT /tag= aj
PT /gene= "megaliI"
PT /function= "AT5"
PT misc_feature
PT /tag= z
PT /gene= "megaliI"
PT /function= "KR3 (inactive)"
PT misc_feature
PT /tag= ak
PT /gene= "megaliII"
PT /function= "KR5"
PT misc_feature
PT /tag= aa
PT /gene= "megaliI"
PT /function= "ACP3"
PT misc_feature
PT /tag= ai
PT /gene= "megaliII"
PT /function= "ACPs"
PT misc_feature
PT /tag= ab
PT /gene= "megaliI"
PT /function= "KS4"
PT misc_feature
PT /tag= am
PT /gene= "megaliII"
PT /function= "KS6"
PT misc_feature
PT /tag= ac
```

```
PT /gene= "megaliI"
PT /function= "AT4"
PT misc_feature
PT /tag= an
PT /gene= "megaliII"
PT /function= "AT6"
PT misc_feature
PT /tag= ad
PT /gene= "megaliI"
PT /function= "DH4"
PT misc_feature
PT /tag= ao
PT /gene= "megaliII"
PT /function= "KR6"
PT misc_feature
PT /tag= ae
PT /gene= "megaliI"
PT /function= "ER4"
PT misc_feature
PT /tag= ap
PT /gene= "megaliII"
PT /function= "ACPe"
PT misc_feature
PT /tag= af
PT /gene= "megaliI"
PT /function= "KR4"
PT misc_feature
PT /tag= aq
PT /gene= "megaliII"
PT /function= "TB"
PT misc_feature
PT /tag= ag
PT /gene= "megaliI"
PT /function= "ACPa"
PT CDS
PT /tag= ah
PT /gene= "megaliII"
PT /product= "megalomicin 6-deoxyerythronolide B synthase 3"
```

```
Query Match 100.0%; Score 600; DB 4; Length 47981;
Best Local Similarity 100.0%; Pred. No. 3,6e-85;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCGCTCCGACGGGTCAACCCCGGGGTGAGTTCAGCTGACCTGACCGCAGTGGTG 60
DB 36307 CCGCGCTCCGACGGGTCAACCCCGGGGTGAGTTCAGCTGACCGCAGTGGTG 36366
QY 61 AGGACGCCCGGTGACCTGCGGCTTACCCGTTCAACGACGAACTTGGCTCCCG 120
DB 36367 AGGACGCCCGGTGACCTGCGGCTTACCCGTTCAACGACGAACTTGGCTCCCG 36426
QY 121 TCCCCCTGGGCGGGTCCCGCAACCGGACGAGTGGGTTACAGCTCGCTGACACC 180
DB 36427 TCCCCCTGGGCGGGTCCCGCAACCGGACGAGTGGGTTACAGCTCGCTGACACC 36486
QY 181 CCGTGCACCTCGGCGGCTCTCCCTGCGCGGACGAGTCTCTGTGTGACCGACGCGCAG 240
DB 36487 CCGTGCACCTCGGCGGCTCTCCCTGCGCGGACGAGTCTCTGTGTGACCGACGCGCAG 36546
QY 241 TACCCCGGCTGACGAGTGTTCGCGACGAGCTGGAACAGCGGGGGCGACCGGTG 300
DB 36547 TACCCCGGCTGACGAGTGTTCGCGACGAGCTGGAACAGCGGGGGCGACCGGTG 36606
QY 301 TGTGTGACCGCGAGTGCAGCGCCCGGATCGGCGCGCACTGCAAGCGCTGACGCGA 360
DB 36607 TGTGTGACCGCGAGTGCAGCGCCCGGATCGGCGCGCACTGCAAGCGCTGACGCGA 36666
QY 361 CCGCCTGTGCACTGTGTCTCTGTCTGCGCGCTGCGAGGGCGGTGCTGTGACGACC 420
DB 36667 CCGCCTGTGCACTGTGTCTCTGTCTGCGCGCTGCGAGGGCGGTGCTGTGACGACC 36726
QY 421 CCAGCCTGACACCTCGGTTGTGTCAGGCGCTGCGGACCGGATGCAAGTCCCC 480
```

Db 36727 CCAAGCTGACACACCTCGCGTTGATCCAGGCGCTCGGCGAGCCGGGATGACGTCCCC 36786  
QY 481 TGTGGCTGTGTGACACAGGAGACGCGCGCGTGAACCGTGGAGACGACGTGATCCGCGCC 540  
Db 36787 TGTGGCTGTGTGACACAGGAGACGCGCGCGTGAACCGTGGAGACGACGTGATCCGCGCC 36846  
QY 541 AGGCATGTGCTGGTGGGCTCGGCGGGTGTGGGCGTGAAGTCCCGCGGTGGGGTG 600  
Db 36847 AGGCATGTGCTGGTGGGCTCGGCGGGTGTGGGCGTGAAGTCCCGCGGTGGGGTG 36906

RESULT 3  
AAQ46806  
ID AAQ46806 standard; DNA; 29879 BP.  
XX  
XX AAQ46806;  
AC  
XX 27-AUG-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 22-DEC-1993 (first entry)  
XX  
XX eryA region of S. erythraea chromosome.  
DE  
XX Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;  
KM erythromycin; condensation; elongation; acyl chain growth;  
KM gene replacement; ss.  
XX  
XX Saccharopolyspora erythraea.  
OS  
XX  
XX Key Location/Qualifiers  
FH 3.10199  
FT CDS /\*tag= a  
FT /\*tag= "ORF 1"  
FT 10218.20921  
FT CDS /\*tag= b  
FT /\*note= "ORF 2"  
FT 20922.29879  
FT CDS /\*tag= c  
FT /\*note= "ORF 3"  
XX  
XX W09313663-A1.  
PN  
XX 22-JUL-1993.  
PD  
XX 17-JAN-1992; 92MO-US000427.  
PF  
XX 17-JAN-1992; 92MO-US000427.  
PR  
XX (ABBO ) ABBOTT LAB.  
PA  
XX Ketz L, Donadio S, Mcalpine JB;  
PI  
XX WPI, 1993-242804/30.  
DR P-PSDB; AAR44430, AAR44431, AAR44432.  
XX  
XX Biosynthesis of specific polyketide analogues esp. erythromycin cpds. -  
PT by introducing altered biosynthetic gene-contg. DNA into microorganisms.  
XX  
XX Claim 27; Fig 2; 133pp; English.  
PS  
XX  
XX This sequence represents a fragment of the Saccaropolyspora erythraea  
CC genome, designated eryA. The polypeptides encoded by this region are  
CC involved in the biosynthesis of the polyketide segment of erythromycin.  
CC eryA is organised in modules and each module takes care of one  
CC condensation step. The precise succession of elongation steps is dictated  
CC by the genetic order of the modules. This fragment may be specifically  
CC altered such that novel polyketide molecules of desired structure are  
CC produced. Three types of alteration may be produced; those inactivating a  
CC single function in a module which does not arrest acyl chain growth;  
CC those inactivating a single function in a module which does affect chain  
CC growth; and those affecting an entire module. The mutations may be  
CC introduced by gene replacement. (Updated on 25-MAR-2003 to correct PN  
CC field.) (Updated on 27-AUG-2003 to correct OS field.)

XX  
SQ Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 U; 0 Other;  
Query Match 47.8%; Score 286.6; DB 2; Length 29879;  
Best Local Similarity 69.7%; Pred. No. 2,2e-36;  
Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;  
QY 1 CCGCGCTCGCGAGAGGTACACCCGGGGGTGAGGTGACCTGAGCGGTACCGCACTGGGTG 60  
Db 23503 CCGCGCTCGCGAGAGGTACACCCGAGGCGTGAAGGCTGAGAGCCCGCTTCGCCG 23562  
QY 61 AGGAGCGCCCGGTGACCTGCGCGGTCTACCGCTTCCAGAGACGAACTTGTGATCCCGG 120  
Db 23563 ACGCGCGCGGTGACCTGCGCGGTCTACCGCTTCCAGAGCGGTACTGCTGCCA 23622  
QY 121 TCCCGCTGGGC---CGGATCCCGACACCGCGAGAGTGGCTTACAGCTCGCTGGC 177  
Db 23623 TCCCGCACCGCGCGGCGCGACGAGGACGAGACGACGCTGCTACAGGTGATGGC 23682  
QY 178 ACCCGCTGACCTCGGCGGTCTCTCGTGGCCGAGCGGATCTGTGTGTGACCGGACCGG 237  
Db 23683 GGGAGCCGAGTGGAGAGCGCTTCGCTGGCCGAGCGCTGTGCTGTGACCGGACCGG 23742  
QY 238 CAGTACCCCGGCTGAGCGGAGGTGTCCGACACGCGCTGGAACAGCGCGGCGGACCG 297  
Db 23743 GCGTGCCTGTCGAGTGTGAGACGCGCATCCGAAGTGACTGAGACAGAGCGGTGCGAG 23802  
QY 298 TCGTGTGTGACACCGCGCACTGCGCGCGCGAGATCGGCGCGCACTGACGCGCTCGACG 357  
Db 23803 TCTGACCTGCGAGAGTGAATCCCGTTGACATCGGACACGCACTGGAAGCGCGGACA 23862  
QY 358 GCAACCGCCCTGTCTCACTGTGATCTCTGTCTCGGCTCGCGAGGCGGTGCTGACAG 417  
Db 23863 CCGACGCTGTGTCACATGTGATGTGCTGTGCTGCCG-----ACGGGAGGCGGTG 23916  
QY 418 ACCCGACCTGAGACACCTTGGCTTGTGTCAGAGGCTCGGCGCGGAGATGAGCGTCC 477  
Db 23917 ATCGTGTGAGACGCGCTGCGCTGTGTGACAGGCTCTGAGAGCGGCGGAGTGAAGAC 23976  
QY 478 CCGTGTGTGTGACACGAGGAGACGCGCGCGCGGTGACCGTGTGAGACGACGTGATCCG 537  
Db 23977 CGCTGTGTGTGACACCGGACGCGCGGTGAGGTGAGCGGAGATGCTGATCCG 24036  
QY 538 CCGAGGCAATGTGTGAGGTGCGCGGGTGTGAGGCTGTGAGTCCCGCGCGGTGGG 597  
Db 24037 CGAGGCAATGTGTGAGGTGCTCGCGCGGTGTGAGCATGACAGACCGGAGCGCTGG 24096  
QY 598 GTG 600  
Db 24097 GCG 24099

RESULT 4  
ADM97245  
ID ADM97245 standard; DNA; 9510 BP.  
XX  
XX ADM97245;  
AC  
XX  
XX 01-JUL-2004 (first entry)  
DT  
XX  
XX Synthetic polyketide synthase gene method related synthetic DBS3 DNA.  
DE  
XX ds; gene; polyketide synthase; PCR; medicine; agriculture;  
KW synthetic gene; DBS3.  
XX  
XX Synthetic.  
OS  
XX W02004029220-A2.  
PN  
XX 08-APR-2004.  
PD  
XX 26-SEP-2003; 2003MO-US030940.  
PF  
XX

PR 26-SEP-2002; 200205-0414085P.  
XX  
XX (KOSA-) KOSAN BIOSCIENCES INC.  
XX  
PI Santi DV, Reid RC, Kodumal SJ, Jayaraj S;  
XX WPI; 2004-316102/29.  
XX  
XX New synthetic genes encoding polyketide synthases, useful for  
PT facilitating production of therapeutic polyketide compounds, or in human  
PT and veterinary medicine, pharmacology, agriculture and molecular biology.  
XX  
XX Example 7, Page 110-112; 172pp; English.  
XX  
XX The present invention relates to a synthetic gene encoding a polypeptide  
CC segment that corresponds to a reference polypeptide segment encoded by a  
CC naturally occurring gene. In particular, the gene is a polyketide  
CC synthase coding sequence. The composition, the gene and methods are useful for  
CC producing synthetic genes or libraries of such genes, and for  
CC manipulating and characterizing genes and corresponding encoded  
CC polypeptides. The synthetic genes may be used for facilitating production  
CC of therapeutic or commercially important polyketide compounds. These may  
CC be used in human and veterinary medicine, pharmacology, agriculture and  
CC molecular biology. The present sequence is a synthetic gene used in the  
CC exemplification of the invention.  
XX  
SQ Sequence 9510 BP; 1587 A; 2635 C; 3236 G; 2052 T; 0 U; 0 Other;  
  
Query Match 31.5%; Score 189; DB 12; Length 9510;  
Best Local Similarity 59.5%; Pred. No. 4.1e-21;  
Matches 359; Conservative 0; Mismatches 235; Indels 9; Gaps 2;  
  
QY 1 CCGCGCTCCGCGAGGCGGTACACCGCGGGGGGAGGAGTGCATGCGCCGACAGTGGTGG 60  
DB 2582 CCGCATTTAGCGAAGCCTATGCGCGCGCGTGAAGGAGTGGTGCATCTTTCGCG 2641  
QY 61 AGGAGCGCCCGGTGACCTGCGCGGTACCGGTTCACAGACAGAACTTGTGCTCCCG 120  
DB 2642 ATGCCCGCCGAGGGAATTACAGTGTATCCGTTTCAGCGCTTACCGCTTACGCTGAGA 2701  
QY 121 TCC---CCTGGGCGGGGTCCCGACACCGGCGACGATGCGCTTACGACCTGCGCTGGC 177  
DB 2702 TTCCGACAGGTGGCGGGGCTGTGACGAAATGATGATGCGCTTACAGGTGTTGGC 2761  
QY 178 ACCCGGCGACCTCGGGCGGTCCCTCCGCGCGGAGGCGCTGCTGCTGATACCGGAGCGG 237  
DB 2762 GTGAAAGCGAATGGAGGTCTGCGTCCCTCGCGCGTGGCTGTGTAACCGGCGCGG 2821  
QY 238 CAGTACCCCGCGCTGAGCGGAGTGGTCCGCGACCGCTGGAACAGCGCGGCGGACCG 297  
DB 2822 GTGTACATCTGAGCTGTCCGATGCCGCTGCAAGGCGCTGAGCAGTCCGGGGGCAACGG 2881  
QY 298 TCGTGTGTGACACCGCGCAGTCCGCGCGCGATCGCGCGCGGACCTGACCGCGTCCAGC 357  
DB 2882 TTTTGACATGCGACGTGAAAGCGTTCACGATCGGACAGGCGGTGGAAGCTGCTGATA 2941  
QY 358 GCACCGCGCTGTGACATGCTGTCTCTGCTCGCGCTCCCGAGCGCGGTCTGTGACG 417  
DB 2942 CTGATCGCTGACGACCGTATGATCGTGTAAACCGGTG-----ATGGGAGAGCTGTG 2995  
QY 418 ACCCGAGCTGACACCTCTCGCGTGTGTCAGGCGCTCGGCGAGCGGAGTGCAGTGC 477  
DB 2996 ATCCGAGTCTGATGCTGTGCTTTGTGAGGCGCTTATGATGCTGTGCTGAGTGCAG 3055  
QY 478 CCGTGTGCTGTGACACCGGAGCGCGCGCGCTGACCGCTGAGAGACGATGATCCG 537  
DB 3056 CGCTGTGCTGTGACACCGGAGTGTGCTGTGCAAGGCTGTGATGATGCTGTGATCTG 3115  
QY 538 CCGAGCGCATGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 597  
DB 3116 CCGAAGCATGTGTGGCGGGCTGGCGCGCTGTGTGTGTAACAACCGGCTGTGCTGTG 3175  
QY 598 GTG 600

DB 3176 GCG 3178  
  
RESULT 5  
AAD55818  
ID AAD55818 standard; DNA; 4725 BP.  
XX  
XX AAD55818;  
AC  
XX 27-OCT-2003 (revised)  
DT 07-AUG-2003 (first entry)  
XX  
XX Micromonospora carbonacea polyketide synthase (PKS) type I gene #8.  
DE  
XX Macrolide; rosamycin; polyketide; polyketide synthase; PKS; enzyme;  
KM gene; de.  
XX  
XX Micromonospora carbonacea.  
OS  
XX  
FH Key Location/Qualifiers  
FT CDS 1..4725  
FT /tag= a  
FT /product= "polyketide synthase"  
XX  
PN CA2391131-A1.  
XX  
PD 19-NOV-2002.  
XX  
XX 26-JUL-2002; 2002CA-02391131.  
PF  
XX 26-JUL-2001; 2001US-0307629P.  
PR  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Yang X, Staffa A, Farnet CM;  
PI WPI; 2003-343556/33.  
XX P-PSDB; AAE37002.  
DR  
XX Novel isolated polypeptide involved in biosynthesis of macrolides by  
PT microorganisms, useful for biosynthesis of macrolides by microorganisms,  
PT preferably for biosynthesis of rosamycin.  
XX  
PS Claim 1; Page 164-166; 206pp; English.  
XX  
XX The invention relates to genes and proteins involved in the biosynthesis  
CC of macrolides by microorganisms. In particular it relates to the nucleic  
CC acids forming the biosynthetic locus for rosamycin (a 16-member  
CC macrolide antibiotic) from Micromonospora carbonacea. The invention is  
CC useful for the biosynthesis of macrolides by microorganisms. It allows  
CC direct manipulation of the proteins involved in the biosynthesis of  
CC rosamycin. It is useful to catalyze certain biochemical reactions, in  
CC vitro or in vivo, to direct or enhance the synthesis or modification of a  
CC polypeptide, polyketide substrate or its precursor. The present sequence  
CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-  
CC OCT-2003 to standardise OS field)  
XX  
SQ Sequence 4725 BP; 482 A; 1782 C; 1885 G; 576 T; 0 U; 0 Other;  
  
Query Match 20.7%; Score 124; DB 10; Length 4725;  
Best Local Similarity 55.0%; Pred. No. 5.9e-11;  
Matches 350; Conservative 0; Mismatches 250; Indels 36; Gaps 4;  
  
QY 1 CCGCGCTCCGCGAGGCGTACACCGGGGGGTGAGGTGACT-----GACGACCGGAG 54  
DB 2642 CTTCCGCGCGCCGAGCGCTTACCCAGGCGGTGAGCGGTGAGTGTGCTGCGGCTTGGC 2701  
QY 55 TGGGTGAGGAGGAGCGCGCGGTGACCTGCGGTCTACCCGTTCAACAGACAGAACTTGTGCG 114  
DB 2702 GGGCGCGCGCGGTGCGCGGTGACCTGCGCGGTGACCGGTTCAGCGGCAACGCTACTGCG 2761

QY 115 TCCCGGTCCCTGAGCGCGGTCTCCCGACA----CGGCGACGAGTGGCGTTACCACTCG 171  
Db 2762 TGAAGCCCGCGCGAGCGGACACCTCCGAGCCCGGAGACGGCTGGCGCTACCGGGTCCG 2821  
QY 172 CTTGGACCCCGTGCACCTCGAGCGGTCTCTCTGCGGACCGAGTCTGTGTATCCG 231  
Db 2822 GCTGGCGCGGCTTTCAGCGGACCGGCGCGCGCGCGGACCGGTGTCTGTGTATCCG 2881  
QY 232 GAGCGGACGATACCCCGGCGCTGACCGGACGTTGTCGCGGACCGGCTGGAACAGCGCGGG 291  
Db 2882 GCCCGAGACGACCGCGGAGAGTGTCAAGGCGGTGCGGACGCGCTCACCGCGCGGGCG 2941  
QY 292 CGACCGGTGTGTGTGTGACCGCGGACGTG-----CGGCGCGGATCGGCGCG 339  
Db 2942 CCGAGGTGGCGGTGTGTATCGGTGACCGGACGACCGACCGGCGCGGTGGCGCGGT 3001  
QY 340 CACTCGACCGCGGTGACGCGGACCGGCGGTGTCTGTGTGTCTGTCTGTG----- 390  
Db 3002 TGCTCACCGGCGCGGTGCGGCGGCGGCGGACCGCGGTGTGTCTGTCTGTGTGTG 3061  
QY 391 -----CGCTCGCGGCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 444  
Db 3062 GTGCGCGGACCGCGGACCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 3121  
QY 445 TCCAGCGGTCTGCGCGGACCGGAGATCGACGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTG 504  
Db 3122 CCGAGCGGTG 3181  
QY 505 CCGCGGTG 564  
Db 3182 TCTCGGTG 3241  
QY 565 GGGTG 600  
Db 3242 GGGTG 3277

RESULT 6  
AAD55810  
ID AAD55810 standard; DNA; 60196 BP.  
XX  
AC AAD55810;  
DT 27-OCT-2003 (revised)  
DT 07-AUG-2003 (first entry)  
XX  
DE Micromonospora carbonacea polyketide synthase (PKS) type I gene cluster.  
XX  
KM Macroliide; rosaramicin; polyketide; polyketide synthase; PKS; enzyme;  
KM gene; ds.  
XX  
OS Micromonospora carbonacea.  
XX  
FH Key Location/Qualifiers  
FT 1..1683  
FT CDS  
FT /\*tag= a  
FT /product= "Polyketide synthase #1"  
FT /note= "CDS does not include start codon"  
FT /partial  
FT CDS  
FT complement(1728..2522)  
FT /\*tag= b  
FT /product= "Polyketide synthase #2"  
FT complement(2629..3861)  
FT /\*tag= c  
FT /product= "Polyketide synthase #3"  
FT 4365..5573  
FT /\*tag= d  
FT /product= "Polyketide synthase #4"  
FT 5702..19117  
FT /\*tag= e  
FT /product= "Polyketide synthase #5"  
FT 19144..24921  
FT CDS  
FT /\*tag= f

FT /product= "Polyketide synthase #8"  
FT /note= "CDS does not include start codon"  
FT /partial  
FT CDS  
FT 2493..36230  
FT /\*tag= g  
FT /product= "Polyketide synthase #7"  
FT 36292..41016  
FT /\*tag= h  
FT /product= "Polyketide synthase #8"  
FT 41049..46403  
FT /\*tag= i  
FT /product= "Polyketide synthase #9"  
FT 46400..47794  
FT /\*tag= j  
FT /product= "Polyketide synthase #10"  
FT /note= "CDS does not include start codon"  
FT /partial  
FT CDS  
FT 47794..49083  
FT /\*tag= k  
FT /product= "Polyketide synthase #11"  
FT 49092..49814  
FT /\*tag= l  
FT /product= "Polyketide synthase #12"  
FT 49868..51226  
FT /\*tag= m  
FT /product= "Polyketide synthase #13"  
FT /transl\_except= (pos:51221..51223, aa:Yaa)  
FT /note= "Yaa corresponds to amino acids from position 452-1811"  
FT CDS  
FT 51506..53416  
FT /\*tag= n  
FT /product= "Polyketide synthase #14"  
FT /note= "CDS does not include start codon"  
FT /partial  
FT CDS  
FT complement(54569..53358)  
FT /\*tag= o  
FT /product= "Polyketide synthase #15"  
FT 54897..56342  
FT /\*tag= p  
FT /product= "Polyketide synthase #16"  
FT 56408..57634  
FT /\*tag= q  
FT /product= "Polyketide synthase #17"  
FT 57657..59123  
FT /\*tag= r  
FT /product= "Polyketide synthase #18"  
FT /note= "CDS does not include start codon"  
FT /partial  
FT CDS  
FT 59363..60196  
FT /\*tag= s  
FT /product= "Polyketide synthase #19"  
XX  
CA2391131-A1.  
XX  
XX  
XX 19-NOV-2002.  
XX  
XX 26-JUL-2002; 2002CA-02391131.  
XX  
XX 26-JUL-2001; 2001US-0307629P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Yang X, Staffa A, Farnet CM;  
XX  
XX WPI; 2003-343556/33.  
XX  
XX P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,  
XX AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,  
XX AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.  
XX  
XX Novel isolated polypeptide involved in biosynthesis of macroliides by  
XX microorganisms, useful for biosynthesis of macroliides by microorganisms,  
XX preferably for biosynthesis of rosaramicin.





CC mystatin PKS is useful as antifungal antibiotics. The present sequence is  
CC a Streptomyces noursei mystatin PKS gene cluster DNA  
XX  
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 U; 0 Other;

Query Match 17.2%; Score 103.2; DB 4; Length 125401;

Best Local Similarity 53.5%; Pred. No. 5.9e-08;  
Matches 291; Conservative 0; Mismatches 238; Indels 15; Gaps 3;

```
QY 72 GTGACCTGCGCGCTACCGCTTCCAGAGAACTTGGGCTCCGGTCCCGCTGGG 131
DB 43840 GGGCAGCAGCTCGCGCGCTGCTCCGACTGTCGCTGGCGCACCCGGCGCAGAC 43899
QY 132 CGGGTCCCGCAGACCGCGGAGAGAGTGGAGTTACAGCTCGCTGCAACCCGCTGACCTC 191
DB 43900 ACATCGGCGACAGACCGCTTCGGGACCGGAGAGTTGGAAACCGCTGCTGCGCGGCG 43959
QY 192 GGGCGGCTCTCTCTCGCGCGGAGCGGCTCTGTGTGTGACCGGAGCG---GCAATACCCCG 248
DB 43960 ACGCGGACACCGCGCGGCTCTGTGTGTGTGCGCCGCGCGGACCAACCGACCCCTGG 44019
QY 249 GCGTGAAGAGAGCTGTGCTCGGAGAGCGCTGGAAGAGCGCGGCGGACCGCTGTGTGTGC 308
DB 44020 GTGCGCGAGCTGTGCTCGCGCGCTGCGCGCGGACCGCGCGGCTGTGTGTGTGTGTGC 44079
QY 309 ACCGCGAGTGTGCGCGCGCGGATCGCGCGCGCTGCAACCGCGCGCTGCAACCGCGCTG 368
DB 44080 GGCACCGAGCGCGCGCGCGCTGCGCGCGCTGTCTACCGAGAGCGCGCGAGACACCGCG 44139
QY 369 TCACATGTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 419
DB 44140 ACCGCGGCTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 44199
QY 420 CCCAGCTGAGACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 479
DB 44200 GGCACCGAGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 44259
QY 480 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 536
DB 44260 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 44319
QY 537 GCCCAGGCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 596
DB 44320 GCCCAGGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 44379
QY 597 GGTG 600
DB 44380 GGTG 44383
```

RESULT 8  
AADS5817  
ID AADS5817 standard; DNA; 11238 BP.

XX AADS5817;

XX 27-OCT-2003 (revised)  
DT 07-AUG-2003 (first entry)

XX Micromonospora carbonacea polyketide synthase (PKS) type I gene #7.

XX Macroliide; rosamycin; polyketide; polyketide synthase; PKS; enzyme;  
KW gene; ds.

XX Micromonospora carbonacea.

XX Key Location/Qualifiers

XX 1..11238

XX /product= "Polyketide synthase"

XX /note= "CDS does not include start codon"

XX /partial

PN CA2391131-A1.  
XX  
PD 19-NOV-2002.

XX 26-JUL-2002; 2002CA-02391131.

XX 26-JUL-2001; 2001US-0307629P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Yang X, Staffa A, Farnet CM;

XX WPI; 2003-343556/33.

XX P-PSDB; AAB37001.

PT Novel isolated polypeptide involved in biosynthesis of macroliides by  
PT microorganisms, useful for biosynthesis of macroliides by microorganisms,  
PT preferably for biosynthesis of rosamycin.

PS Claim 1; Page 151-158; 206pp; English.

CC The invention relates to genes and proteins involved in the biosynthesis  
CC of macroliides by microorganisms. In particular it relates to the nucleic  
CC acids forming the biosynthetic locus for rosamycin (a 16-member  
CC macroliide antibiotic) from Micromonospora carbonacea. The invention is  
CC useful for the biosynthesis of macroliides by microorganisms. It allows  
CC direct manipulation of macroliides and related chemical structures by  
CC chemical engineering of the proteins involved in the biosynthesis of  
CC rosamycin. It is useful to catalyse certain biochemical reactions, in  
CC vitro or in vivo, to direct or enhance the synthesis or modification of a  
CC polyketide, polyketide substrate or its precursor. The present sequence  
CC is M. carbonacea polyketide synthase (PKS) type I gene. (Updated on 27-  
CC OCT-2003 to standardise OS field)

SQ Sequence 11238 BP; 1110 A; 3596 C; 4707 G; 1825 T; 0 U; 0 Other;

Query Match 16.7%; Score 100.2; DB 10; Length 11238;

Best Local Similarity 52.9%; Pred. No. 2.6e-07;  
Matches 250; Conservative 0; Mismatches 208; Indels 15; Gaps 1;

```
QY 143 CACCGGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 202
DB 2249 CACCTGAAACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3008
QY 203 CCGCGCGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 262
DB 3009 GGTCCCGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3068
QY 263 GGTCCCGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 322
DB 3069 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3128
QY 323 CGCCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 382
DB 3129 GCTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3188
QY 383 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 428
DB 3189 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3248
QY 429 -GACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 487
DB 3249 CGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3308
QY 488 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 547
DB 3309 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3368
QY 548 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
DB 3369 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3421
```

```
RESULT 9
AD251725/c
ID AD251725 standard; DNA; 138203 BP.
XX
XX AD251725;
XX
XX 14-UTL-2005 (first entry)
XX
XX FR-008 polyketide gene cluster, SEQ ID 1.
XX
XX Antifungal; antiparasitic; polyketide; gene; ds.
XX
XX Streptomyces sp.; FR-008.
XX
XX Key location/Qualifiers
XX FT complement(574. .1950)
XX FT /tag= a
XX FT /product= "FAD-dependent monooxygenase, FscO, SEQ ID 20"
XX FT complement(2264. .3037)
XX FT /tag= b
XX FT /product= "ADC lyase, PabC, SEQ ID 22"
XX FT complement(3150. .3818)
XX FT /tag= c
XX FT /product= "Regulator protein, FscRI, SEQ ID 10"
XX FT complement(4377. .7205)
XX FT /tag= d
XX FT /product= "Regulator protein, FscRII, SEQ ID 11"
XX FT complement(7210. .10320)
XX FT /tag= e
XX FT /product= "Regulator protein, FscRIII, SEQ ID 12"
XX FT complement(10298. .13315)
XX FT /tag= f
XX FT /product= "Regulator protein, FscRIV, SEQ ID 13"
XX FT 13522. .14898
XX FT /tag= g
XX FT /product= "Glycosyltransferase, FscMI, SEQ ID 17"
XX FT 14953. .16011
XX FT /tag= h
XX FT /product= "GDP-ketosugar aminotransferase, FscMII, SEQ ID
XX FT 16106. .17287
XX FT /tag= i
XX FT /product= "Cytochrome P-450 monooxygenase, FscP, SEQ ID
XX FT 17334. .17528
XX FT /tag= j
XX FT /product= "Ferredoxin protein, FscFE, SEQ ID 15"
XX FT 17556. .18413
XX FT /tag= k
XX FT /product= "Thioesterase, FscTE, SEQ ID 16"
XX FT 18610. .20781
XX FT /tag= l
XX FT /product= "ADC synthase, PabAB, SEQ ID 21"
XX FT 20927. .26158
XX FT /tag= m
XX FT /product= "PKS FscA, SEQ ID 2"
XX FT 26333. .27340
XX FT /tag= n
XX FT /product= "ABC transporter, FscTI, SEQ ID 8"
XX FT 27561. .28280
XX FT /tag= o
XX FT /product= "ABC transporter, FscTII, SEQ ID 9"
XX FT complement(28983. .60860)
XX FT /tag= p
XX FT /product= "PKS FscC, SEQ ID 3"
XX FT complement(60962. .77587)
XX FT /tag= q
XX FT /product= "PKS FscB, SEQ ID 4"
XX FT complement(77983. .84132)
XX FT /tag= r
XX FT /product= "PKS FscF, SEQ ID 5"
XX FT complement(84170. .107485)
XX FT /tag= s
XX FT CDS
```

```
FT /product= "PKS FscE, SEQ ID 6"
FT complement(107496. .136148)
FT /tag= t
FT /product= "PKS FscD, SEQ ID 7"
FT complement(136558. .137766)
FT /tag= u
FT /product= "GDP-mannose-4,6-dehydratase, FscMIII, SEQ ID
XX FT 19"
XX
XX US2005089884-A1.
XX
XX 28-APR-2005.
XX
XX 06-APR-2004; 2004US-00819386.
XX
XX 23-OCT-2003; 2003KR-00074035.
XX
XX (KOAD ) KOREA ADV INST SCI & TECHNOLOGY.
XX (UYSH-) UNIV SHANGHAI JIAOTONG.
XX
XX Lee SY, Deng Z, Chen S, Jeong KJ, Zhou X;
XX
XX WPI, 2005-322033/33.
XX
XX P-PSDB; AD251726, AD251727, AD251728, AD251729, AD251730, AD251731,
XX AD251732, AD251733, AD251735, AD251736, AD251737, AD251738,
XX AD251739, AD251740, AD251741, AD251742, AD251743, AD251744, AD251745,
XX AD251746.
XX
XX New gene cluster for the biosynthesis of FR-008 polyketides, useful for
XX increasing the productivity of FR-008 polyketides or to produce new FR-
XX 008 polyketide variants.
XX
XX claim 1; SEQ ID NO 1; 22pp; English.
XX
XX The present invention relates to a gene cluster (AD251725) for the
XX biosynthesis of FR-008 polyketides derived from Streptomyces sp. FR-008.
XX FR-008 polyketide, a heptene macrolide having aglycone containing 4-
XX aminocetophenone, has antifungal activity and also high toxicity against
XX mosquito larvae, and so is highly useful in agricultural and medical
XX fields. The gene cluster comprises the following genes: fscA, fscC, fscB,
XX fscF, fscE and fscD, which code for modular polyketide synthase (PKS);
XX fscRI and fscRII, which code for ABC transporter proteins; fscRI, fscRII,
XX fscRIII and fscRIV genes which code for regulator proteins; fscP which
XX codes for cytochrome P450 monooxygenase; fscTE which codes for ferredoxin
XX protein; fscTB which codes for thioesterase; fscMI which codes for
XX glycosyltransferase; fscMII which codes for GDP-mannose-4,6-dehydratase;
XX aminotransferase; fscMIII which codes for GDP-mannose-4,6-dehydratase;
XX fscO which codes for FAD-dependent monooxygenase; PabAB which codes for 4
XX -amino-4-deoxychorismate (ADC) synthase and pabC which codes for ADC
XX lyase. The genes of the gene cluster of the invention can be used to
XX develop recombinant microorganisms capable of producing FR-008
XX polyketides, also to increase the productivity of the existing FR-008
XX polyketides or to produce new FR-008 polyketide variants, by its
XX modification.
XX
XX Sequence 138203 BP; 18153 A; 48202 C; 55342 G; 16506 T; 0 U; 0 Other;
XX
XX Query Match 16.7%; Score 100.2; DB 14; Length 138203;
XX Best Local Similarity 53.7%; Pred. No. 1.7e-07;
XX Matches 341; Conservative 0; Mismatches 258; Indels 36; Gaps 5;
XX
XX 2 CGCGCTCCGCGGAGGCTACACCGGGGGGTGAGTTCAGCTGACCGGACGTGGGTGA 61
XX DB 92504 CCCGCTTCAGACGGCGCTTCTTGACCGCCGCTTCAGAGCGGAGACTGACCGCTTCGCCAC 92445
XX
XX 62 GGAGACGCGCGCGGTGACCTGCGCGGTCTACCGCTTTCACAGACGAACTTCTGCGTCCGGGT 121
XX DB 92444 CGACCTGCGCGGTGACACCGAGGCGCTGCGCGGTATCTCCGGCCCTTTCACACTCTGGCG 92385
XX
XX 122 CCCCTGAGCGCGGTGCGCGGACCGGCGAGCGTGGCGCTTACCGAGCTGCGCA--C 179
XX QY 92384 GCGCGGCGCGCGGAGCAGGCGCATGTGATCCAAACGCGACACGAGACGTTGAGAGCC 92325
XX
```

Accession	Gene	Location/Qualifiers
Oy	180	CCGGTGCACCCGAGGAGGATGCTCTCCCTGAGCCGAGACGGATCTG-----GTGGTACACGG 232
Db	92324	GCTGTCCGTGCGCCGCCGATGCGCCCGCCCGCCGCGGAACTTGACTGAGCGGTGTGCTCCCG 92285
Oy	233	AGCGGCAAGTACCCCGGCGCTGAGACGACGATGTCGCGACACGCGCTGGAACAGCGCGAGGC 292
Db	92264	CGCGTACGAGGAGAACCCCTTGAGACGCGCGCGTCTTGACGCGCTGCGGACCGACGATCGT 92205
Oy	293	GACCGTGTGTTGTGCACCGGCGAGTGGCGCGCCCGGATGAGCGCGCATCTGACG---C 349
Db	92204	CCGCGCTACCGGTGACACCGTTCGAGCGCGACAGCGCGCTCGCGACCGCGCTGCGCGCTTC 92145
Oy	350	CGTGCACGAGCAACGCGCCCTGTGCACTGATGATCTCTCTGCTGCGGCTGACCGA----- 400
Db	92144	CGCGGACGAGCGCGCGCCCTTCGCTCGGTGTGATCTCTCTGCTGAGCCCTTGCGGACACGCGG 92085
Oy	401	-----GGGCGGATGCTGTTCGACGACCCCGAGCTGGAACACCTCGCTGTGTGTCGA 448
Db	92084	GGCAGCGGTCTCGCGGCGACCCGAGGAGAGCCCCCGCGCGACCGCGCTGTGTTCA 92025
Oy	449	GCGCGTCGAGCGAGCCCGGATCGACGTTCCCTGTGTGCTGTGACGAGGACCGCGCGC 508
Db	92024	GAGCGGTGTCGACCCCGGAGTCGACGCGCCCTGTGTGCTGTGACCGCGGAGCCGTTCG 91965
Oy	509	CGTACCGCT---CGGAGACGACGATCGATCCGCGCGAGGACCATGTGCGTGGGCTCGGCGC 565
Db	91964	GATGCGGAGTTCCGAGGCGCGTACCGCGCCCGCGCGAGCGCGCGCTGTGGGCTGTGGGCG 91905
Oy	566	GATGATGAGCGTGAAGTCCCGCGCGCGATGGG 600
Db	91904	GATGCGCGCGCTTCGAAACACCCCGCGATGGGCG 91870
RESULT 10		
AAV21187		
ID	AAV21187	standard; DNA; 53789 BP.
AC	AAV21187;	
DT	24-JUL-1998	(first entry)
XX	Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.	
XX	Amycolatopsis mediterranei rifamycin; synthesis; gene cluster;	
KW	polyketide synthase; actinomycete; ansamycin; ds.	
XX	Amycolatopsis mediterranei.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1825..15543
FT		/*tag= a
FT		/label= ORF_A
FT		/product= "polyketide synthase"
FT	CDS	15550..30759
FT		/*tag= b
FT		/label= ORF_B
FT		/product= "polyketide synthase"
FT	CDS	30895..36060
FT		/*tag= c
FT		/label= ORF_C
FT		/product= "polyketide synthase"
FT	CDS	36259..41325
FT		/*tag= d
FT		/label= ORF_D
FT		/product= "polyketide synthase"
FT	CDS	41373..51614
FT		/*tag= e
FT		/label= ORF_E
FT		/product= "polyketide synthase"
FT	CDS	51713..5293
FT		/*tag= f
FT		/label= ORF_F
FT		/product= "polyketide synthase"

Query Match	Best Local Similarity	Score	DB 2	Length	53789
Matches 256; Conservative	0	Mismatches 220	Indels 6	Gaps 2	
125	CCCTGGGCGCGGTC	CCCGACACCGGCGACGAGTGGCGTTACCAAGCTCGCTGGCA	CCCCGT	184	
13599	CCGCGCCCGCGAGCGT	CGACCGCGGAGAAAGTCGCGTACCAAGTCACTGGCAGCCCT		13658	
185	CGACTCGCGGCGGT	CCTCCTGGCGCGACCGGTCGTGTGTGACCGGAGCGGCAATAC		244	
13659	GGAGCGCGAAAGCGCGCGCGTGC	CGGCGCGGCGCGTGCCTGCTGCGTCCGCGCGGCGAC		13718	
245	CCCGCGCTGACCGACG	GTGTCCGCGACCGGCGCTGGAAACAAGCGGCGACCGTCTGT		304	
13719	CAACGACGCGCTCT	GAAAGAGTGAACCGGCGCAAGGACTGACATCGTCCGCTGGAGAT		13778	
305	GTGACACGCGCAGT	CGCGCGCGGATCGGCGCGGACCTGACACGCGCGTGAACCGCACCGC		364	
13779	CGAGAAAGCTT	CGCGGACACGCTCGCGGACGAGCTGCGGAAAGCTCTGGGAGGACAGA		13838	
365	CCGTGTCACGTGTGTCT	CTGTGCTCGCGCTCGCCGAGGCGGTGCTGTC--GACGACCC		421	
13839	CTTCACCGGCGTGTCTGT	CGCTGTGCTGTGCTCTGACACGCGCGCCCGCGGACGCGCGGAGAT		13898	
422	CAGCTGGAACCT	CGCGCTTGTGTGACGCGGCGCTGGGCGACCGCGGATCGACGTCCCGCT		481	
13899	CAACCGGTGACCT	CGCGGTGTGTGACGCGCGCTGGGCGACCAACACGTCGCGCGCT		13958	
482	GTGCGTGTGACCA	CGGAGACCGCGCGCGTGAACCGTGGAGACGACGT---GATCCGAC		538	
13959	GTGTGTCTCACTT	CGCGCGCGGTGAACATCGGATCAAGACCGCGTGAACCGCACCGGC		14018	







CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
CC biologically active variant, where the nucleic acid sequence is not  
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or ~  
CC Streptomyces antibioticus. The invention also relates to a macrolide  
CC biosynthetic gene cluster, or fragments thereof. The macrolide  
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
CC pikromycin, neomethymycin, narboxmycin or a combination of these  
CC compounds. Recombinant or augmented cells comprising the desosamine  
CC and/or macrolide biosynthetic gene clusters are useful for the production  
CC of biologically active macroclides. The macroide biosynthetic proteins  
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and  
CC narboxmycin. The alternative termination of polyketide synthesis may be  
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
CC monomers. The compounds produced by the recombinant host cells are useful  
CC as biopolymers, e.g., in packaging or biomedical applications, to  
CC engineer PHA monomer syntheses or to prepare biologically active agents,  
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
CC chronic obstructive pulmonary disease as well as other diseases involving  
CC respiratory inflammation, cholesterol-lowering agents or macroclide-based  
CC antibiotics which are active against a variety of organisms, e.g.,  
CC bacteria, including multi-drug resistant pneumococci and other  
CC respiratory pathogens, as well as viral paracellular pathogens, or as crop  
CC protection agents (e.g., fungicides or insecticides) via expression of  
CC polypeptides in plants. The present sequence represents the macrolide  
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439,  
XX as given in figure 31. (Updated on 15-SEP-2003 to standardise OS field)

Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 0 U; 2 Other;



CC derived from the eryC gene cluster of *Saccharopolyspora erythroa* or  
CC Streptomycetes antibiotics. The invention also relates to a macrolide  
CC biosynthetic gene cluster, or fragments thereof. The macrolide  
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
CC plitromycin, neomethymycin, narbomycin or a combination of these  
CC compounds. Recombinant or augmented cells comprising the deoxamine  
CC and/or macrolide biosynthetic gene clusters are useful for the production  
CC of biologically active macrocides. The macroide biosynthetic proteins  
CC are useful for synthesis of methymycin, plitromycin, neomethymycin and  
CC narbomycin. The alternative termination of polyketide synthesis may be  
CC useful to prepare novel antibiotics and polyhydrotrolykanoate (PHA)  
CC monomers. The compounds produced by the recombinant host cells are useful  
CC as biopolymers, e.g., in packaging or biomedical applications, to  
CC engineer PHA monomer synthases or to prepare biologically active agents,  
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
CC chronic obstructive pulmonary disease as well as other diseases involving  
CC respiratory inflammation, cholesterol-lowering agents or macroide-based  
CC antibiotics which are active against a variety of organisms, e.g.,  
CC bacteria, including multi-drug resistant pneumococci and other  
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
CC protection agents (e.g., fungicides or insecticides) via expression of  
CC polyketides in plants. The present sequence represents the macroide  
CC biosynthetic gene cluster (pit) from *Streptomyces venezuelae* ATCC 15429.  
CC (updated on 15-SEP-2003 to standardise OS field)

XX  
XX  
SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 U; 0 Other;

Query Match 14.9%; Score 89.4; DB 3; Length 37948;  
Beef Local Similarity 52.6%; Pred. NO. 1e-05;

QY	1	CCGGGCTCGCCGAGCGGTACACCCGGGGGTGAGAGTCCACTGGCGTACCCGATGG---G	57
Db	8760	CCTGCGCTCGCGAGCGGTGGGTCAACGGGCTTCCCGTGGCATGCACTTGGCTCCTGCCCC	8819

Db 8820 CCACGCGCTCTCCGCCGCCCGGTCTGCCCCACCTACGCGTTCCAGGCGGAGCGCTACTGAGCTCG 8819

1

Qy 178 ACCCGATGACCTCGAGGCGGTCTATCCCTGAGCCGAGACGAGTCTGTGGTGTGACCGAGAGCGG 237  
 Db 8937 AGCGCTTCCGAGCGGCGGAGGAGGTCTGAGCGGACCGG---CTGTATCCGAGCGTATGCTCTCG 8993

[illegible][illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

0.17

[illegible]

**TGAGGAACCGAATTC**

100

594

100

**B**

$$=$$

Mon Apr 10 09:27:47 2006

us-10-611-442-2\_copy\_1\_600.rng

Page 15

Db 9351 GGG 9353

Search completed: April 7, 2006, 03:16:47  
Job time : 346.081 secs

---

**This Page Blank (uspto)**

strains include: 1) 11 isolates from clinical samples of

## ORIGIN

Pile 'C', Pile 'D', Pile 'S', Pile 'P', M18851, M18858, M27912, M27931, and M28869: 2) 1 strain from American Type Culture Collection (ATCC): 27853 "

Query Match 2.8%; Score 253.2; DB 10; Length 1521;

Best Local Similarity 53.7%; Pred. No. 4.8e-35;

Matches 575; Conservative 0; Mismatches 483; Indels 12; Gaps 2;

```
QY 3949 TGGTGGCGAGTCGATACGTGCAACCCGCGGATGTCTCTCATGTGACACCGAGGGC 4008
DB 1077 TGGGCGCGGAAGCATGAGCGCGCGACGAACTGAGCTGCGCGCGCGCAT 1018
QY 4009 CGGTGCGGAGCGGACGACCCGCTTCCGCGAGCCACTGCTGCTC-----TGCTCAGCC 4059
DB 1017 AGATCGGACGACGACGACCCGCGAGAAATGACCCGCGACCGGAAAGATTGGCGGCG 958
QY 4060 GCAACGGTGGCTCCGTGCGGTAGGGCGTCACTGTGTGACAGCGCGGAGAAATGAGGCT 4119
DB 957 CCGCGGAGGCGAGGAGATGATACATGGGAAAGTGGTATGATGAGTGAAGAGAGCGC 898
QY 4120 GTGCGACGACCTTCTCCGCGCGGACGATGCGCTGACGCGGTCAACGCTGATGCGCGTGG 4179
DB 897 GCACACGACATCCCTGTGCGCGAGCGCTGTGTAAGCTCGTCCGCGGAAAGC---GGA 841
QY 4180 CGGTGCGGTCAACGAGATGATCACTGTGTGTGTCTCTCCGTGCGCGGAGCG 4239
DB 840 AATGCTCTTCCACGAGATCGGAAATGAGCGTGTGTGCGAAGTCTGGTACCAAGCC 781
QY 4240 AGTCACGAGTACGCGCGGTACGTCGCGACACTGCTGTGTAGAGCGCGGTGTGACCC 4299
DB 780 GGAAGGACATGATGACAGGAGCGCGCGCACAGCGCGCGATGAGCGCTCGCGCATCTCC 721
QY 4300 GGTGTGTCTCCGTGTCTCGCGGACCGGTGAGAGAGTAAAGCCATGCGCGCGCGC 4359
DB 720 TGCGGTGCTCCAGGCGCTGTGTCAGCTGTGATGAGAGCGCGGACCGACGCGCGA 661
QY 4360 ACTGCTCATCTGTCGCTGTGTGTCGCTGTGATGACCACTGTGCGCGCGGATGCGCA 4419
DB 660 CCGGTGATGCTTGGCGGTGATGCGCGGACATGACCACTGTGTCGCGGAGCGCA 601
QY 4420 AGTGTGATGAGCGCGGATCGCTTTCGCGACAGCGCGTGTGATGACCAAGCGCGC 4479
DB 600 AATCTTAAAGTGTGATGAGCGGTGCGCGGTGTTTCTCGAGGATGATGCGAGCGC 541
QY 4480 CCGTGAAGCGGATGACCGCTTGTGTGCGGTGAGTGAACCTGCGCGCTGACCGGATC 4539
DB 540 CCGTGAAGGTGTGAAACCTTGTGTGCGGTGAGTGAAGTGTGCGCGCTGCGCGTGC 481
QY 4540 CGCCGACGAGTCAACCGCTGTGTCAGCGCGCGGCGCGCGGTGCGGAGAGAGCT 4599
DB 480 GCAAGACCTGTGCGCGGTGTCGCGACGCGGAGAGAGTGTGCGCGCATGTGATGACT 421
QY 4600 TGAAGTGTGTGCGCGGATCTTTCGCGAGCGCTTCCACAGTGTGCTGTGCGCGCA 4659
DB 420 TCAAGTGTGTGATGATGCGGATCGCGTGTGATGCTTCCACAGGAGGTGCGGTAGC 361
QY 4660 GGTGTACGCGGATGCGCGCGGTGTGCGGAGTGAACAGCGCGGACGAGTGTGCGGT 4719
DB 360 AATGACGCGGAGATGCGGTGTGTGCGGAGTGTGCGCGCTGATGCTGCGCGAGCGGT 301
QY 4720 CGACGACGCGTCCGCGGTGTGATGCGGAGAAACAGGATGTGATGCGGACGAGCTG 4779
DB 300 CGAAGATTCAATGAGTCTGTATGATGTCCACGAAATGCGCTTGAAGCCGTTTCCACAG 241
QY 4780 CCGGTGCGCGGTGCGCGGAGTCAATTGACGCGATGATCACTTCAACGCGTGAAGTAC 4839
DB 240 GCGGTGTGTGTGCGGAGAGATGAGCGGTGTGATGATGCTGCGCGGTGACCGGAG 181
QY 4840 CGGCGCGGACGACGATTCAGAGGCGAGCGGTGCGGTGCGAGTGTGCGAGTGTGCGGT 4899
DB 180 ATGCTGTGAGCGCGGTGATGAGCGCGAGGTGTGCGGTGCGGAGAGACGATGTCTTCA 121
```

```
QY 4900 CCCGACGAGTGTGCGGACCGCGGCTCGAATCCCGACACGAGGCGCGCGGTGTGTA 4959
DB 120 CCGCGAGTATGCGACAGGAGCTCTTCCACGATGAGTGTGATGCGCGCGCGGTGTGTA 61
QY 4960 GCCAGTGTGTGTCAGAGGCGCACTCCAGCGCGGCGCGGAGAACCGCTGCCG 5009
DB 60 GGAAGGCGGTCTCCAGATCTTTCAGGTGAGGAGATGATGATCTTCCAGG 11
```

## RESULT 2

BZ58756/c 1135 bp DNA linear GSS 17-DEC-2002

LOCUS P98401.457.s1 pac82-164 Pseudomonas aeruginosa genomic clone

DEFINITION P98401.457, genomic survey sequence.

ACCESSION BZ58756

VERSION BZ58756.1 GI:27173859

## KEYWORDS

ORGANISM

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1135)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, B.E., Hastings, M.,

Burns, J.L., Kaul, R., and Olsen, M.V.

Whole-Genome-Sequence Variation among Multiple Isolates of

Pseudomonas aeruginosa Library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..1135

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="P98401.457"

/clone\_1db="pac82-164"

/note="clinical isolate 2-164 Whole genomic shotgun library."

## ORIGIN

Query Match 2.3%; Score 207.8; DB 9; Length 1135;

Best Local Similarity 54.0%; Pred. No. 6.7e-27;

Matches 447; Conservative 0; Mismatches 378; Indels 3; Gaps 1;

```
QY 4050 TGTGTACGCGGACGAGTGTGCTCGTGTGAGGCGGTATCTGTGACAGCGGAGAG 4109
DB 873 TTTCGCGGACCGCGGAGGACGAAATGTATCATGAGAGTGTGCTTATAGTGTGAG 814
QY 4110 AAGTACGCTGTGACGACCTTCTCCGCGCGGACGATGCGCTGACGCGGTGACGCTG 4169
DB 813 AAGTACGCGCTTACGAAATCTTGTAGGCGAGGCGGTGATACAGTGTGCGCGGAA 754
QY 4170 ATGCGGTGCGGTGCGCTTCCACGAGATGATCACTGTGTGTGTCTCTCTGCTG 4229
DB 753 A---GCGTAAATGCTTTCACACGAGTCCGGAATGAGCGTGTGTGCGGAGGTG 697
QY 4230 GCGGAGCGGATGACGAGTACCGCGGTACGTCGCGGACGCTGCTGTGTGAGCGCG 4289
DB 696 GTACAGGCGCGGACCTGTATGTCAGAGAGCGCGGACGAGAGCGCGATGAGCGCTG 637
QY 4290 TGTGTACCGGTTGTGCTCTGCTGCTGCGCGGACGCGTGTGAGGAGTGAAGCCCATG 4349
DB 636 GCGATCTCAATGCGGTCTTCCAGGCTTGTGTGACGCTGTGATGAGGAGCGCGGAG 577
QY 4350 GCGCGGCGCACTGCTCATCTTTCGCTGTGTGTCGATCTGAGTGTGAGCACTTGTGCGG 4409
```

Db 576 GCAGCCCGACCTCGTTGAGCTTCCGCTTGATGCGGGCATGACCAACCTGACTCTCG 517  
Qy 4410 CCGATGCCGAAGTTGTCATGCGCCGATCCGTTCCGCGACAGGCCCTGCTGATACC 4469  
Db 516 ACBAAGCCCAACTCTTAAGTGTGAGCGTTCCGGGTGTTTGTCCAGACAGATG 457  
Qy 4470 ACCGCCCGCCCTGAAAGCGGTGACCGCTTGTCGTCGTAAGTGAACACTCCGCG 4529  
Db 456 ATCGCACCGCCCTGAAAGGTGTGAACACTTGTTGTCATGAAAGTGAAGATGCTAGG 397  
Qy 4530 TCACCGGATCCGCCACCGACGTCACCCGTGTGACAGCGGCGTGGCGCGCTCG 4589  
Db 396 TCCCCTGTGGCGACACGCTGCGCCGCTCCGCGACCGCGAAGCATCGCGCATCG 337  
Qy 4590 AAGAAGCTTGAAGCTGTGTGTGGGGCGGATCTTGCCGACCCCTCCAAAGTGTGCT 4649  
Db 336 TGAATGACCTTCAAGTGTGATATATTCGGGATCCGCTGATCCCTCGAATCGCAGGG 277  
Qy 4650 CCGCCCGACAGGTGTACGCCGACATCGCCCGGTCTGCGGGGTGACAGACGCGCGACG 4709  
Db 276 TTGCGGTAGGAATGACACCGGACAGATCGCGGTGTCTGCGGGGTGATCGCGCTGATC 217  
Qy 4710 TGGTCCGGGTCCACCAAGACCGGTGCGCGGTGCACTGCGAACAACCGGTGTGATGCTG 4769  
Db 216 CGGACGGGTGAGATTCAATGAGTCTGTGTCATGTCACGAAGATCGGCTTGAAGCCG 157  
Qy 4770 AGCCAGCTGCGCGCGTGGCGCGGTGCGCGAGTCAATGACCGCATGACTTCAACCG 4829  
Db 156 TTTCACAGACAGCGCGTGGCTGTGTGCGACGAAGAGTACCGCTGTGATGACACTCGCG 97  
Qy 4830 GTGACGTCAACCGCGCCGACCAAGTTCAGGGCGACCGGTGCGCTTG 4877  
Db 96 GTACCCCGCAATGCTGACGCGCGGTGATGAGCGCCAGGGTGGCTTG 49

RESULT 3  
BZ298798/c 584 bp DNA linear GSS 31-OCT-2002  
LOCUS CG4587.r1 Candida glabrata Random Genomic Library Candida glabrata  
DEFINITION genomic clone CG4587, genomic survey sequence.  
ACCESSION BZ298798  
VERSION BZ298798.1 GI:24442259  
KEYWORDS GSS.  
SOURCE Candida glabrata  
ORGANISM Candida glabrata  
Bukhryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
REFERENCE Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.  
AUTHORS 1 (bases 1 to 584)  
TITLE the 'asexual' pathogenic yeasts Candida glabrata  
JOURNAL Genome Biol. 4 (2), R10 (2003)  
PUBMED 12620120  
COMMENT Contact: Wong S  
Department of Genetics, Smurfit Institute  
Trinity College Dublin  
Dublin 2, Ireland  
Tel: 353 1 6082319  
Fax: 353 1 6798558  
Email: swong@cd.ie  
Class: plasmid ends.  
FEATURES  
Location/Qualifiers  
1..584  
/organism="Candida glabrata"  
/mol\_type="genomic DNA"  
/strain="CBS 138"  
/db\_xref="taxon:5478"  
/clone="CG4587"  
/clone\_1ib="Candida glabrata Random Genomic Library"

## ORIGIN

Query Match 1..6%; Score 141.6; DB 9; Length 584;  
Best Local Similarity 56.2%; Pred. No. 5.2e-15;

Matches 289; Conservative 0; Mismatches 219; Indels 6; Gaps 1;  
Qy 4411 CAAATGCCAAGTTGTCATGAGCCCGATCCGTTGCGCACAGCCGCTGCTGATCCA 4470  
Db 557 CCAATGCCCTGTGTGCGACACAGCCGACCGCTTGGCGGACACTGTGCTGCTGCGACACA 498  
Qy 4471 CCGCCCGCCCTCGAAGCGGTGACCGCTTGTCGTCGTAAGTGAACACTCCGCGCT 4530  
Db 497 CCAATGCCCTGTGCGGACGATCATTTCTTCTCGGTGAGGCTGAACATGCGCAACG 438  
Qy 4531 CACCGGATCCGCCACCGACGTCACCCGTGTGACGCCCAAGGCGTGGCGCGCTCGA 4590  
Db 437 TGCAGAACAGACCGACCGACCGCGCTCGACGCTGCGCGCTGCGCGCGCTCT 378  
Qy 4591 AAGAAGCTTGAAGCTGTGTGTGGCGGATCTTCCGACGCTTCCACAGTGTGCTG 4650  
Db 377 CGAAGACAGACAGACGCTCTCTCGGCAATGCTCCGAGCGCGCTCATGTCCGCGGCT 318  
Qy 4651 GGCCCACAGGTGTACCGCGACGATGCGCGCGCTCTGCGGGGTGACAGAGCGCGGACGT 4710  
Db 317 GACCTGAGAGGTGACACCGCATGATGCTCTTGTGCGCTCGGTGATGCGCGCGACCG 258  
Qy 4711 GGTCCGGGTGACACAGACCGGTGCGCGGTGCACTGCGAACAACCGGTGTGATCCGA 4770  
Db 257 CGTCCGGCGACAGGACGAAGTGTGCTCGCTCGATGTCCGGAACAACCGCGTCCGCG 198  
Qy 4771 GCCAGTCCCGCGGTGCGCGGTGCGCGGACGATGATGATGATGATGATGATGATGATG 4830  
Db 197 TCAGCGGACCAAGTGTGCGGTGCGCGGACGATGATGATGATGATGATGATGATGATG 138  
Qy 4831 T-----GACGTACACCGCGCGCGGACGATGATGATGATGATGATGATGATGATG 4884  
Db 137 CCGGCGCATGCTCGCGGAGAGACCAAGTGTGACCGAGAGTCCCGAGATTACCG 78  
Qy 4885 CGATGCAATGCCGTACCCCGACCAAGTTCGCGCAC 4918  
Db 77 CGACGCAAGTGGCGCCCGCGACGACGATCGGCG 44

RESULT 4  
BZ562892 719 bp DNA linear GSS 17-DEC-2002  
LOCUS pac62-164\_3982.x1 pac62-164 Pseudomonas aeruginosa genomic clone  
DEFINITION pac62-164\_3982, genomic survey sequence.  
ACCESSION BZ562892  
VERSION BZ562892.1 GI:27185631  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE Spencer,D.H., Raymond,C.K., Smith,E.B., Sims,E.E., Hastings,M.,  
AUTHORS 1 (bases 1 to 719)  
TITLE Whole-genome-sequence variation among multiple isolates of  
JOURNAL Pseudomonas aeruginosa library  
P. Bacteriol. (2002) In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.  
FEATURES  
Location/Qualifiers  
1..719  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pac62-164\_3982"  
/clone\_1ib="pac62-164"

ORIGIN  
/note="clinical isolate 2-164 Whole genomic shotgun library."

Query Match 1.44; Score 127.2; DB 9; Length 719;  
Best Local Similarity 56.0%; Pred. No. 2e-12;  
Matches 262; Conservative 0; Mismatches 203; Indels 3; Gaps 1;

Qy 4410 CCGATGCGGAAGTGTGATGAGCGCGGATCGTTCGCGCAGCAGCGCGCGTGTGACG 4469  
Db 168 CCGAGCGCGTGAAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 227  
Qy 4470 ACCG 4529  
Db 228 AACATCG 287  
Qy 4530 TCACCGGATCG 4589  
Db 288 CCGCGCGGGA---GCCCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344  
Qy 4590 AAGAAGAGCTTGACCTGTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4649  
Db 345 TCAGCGAGGCGAGTGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 404  
Qy 4650 CCG 4709  
Db 405 GAGCG 464  
Qy 4710 TGTGCG 4769  
Db 465 GGTGCG 524  
Qy 4770 AGCG 4829  
Db 525 AGCG 584  
Qy 4830 GTGACG 4877  
Db 585 GCGAGCG 632

RESULT 5  
CG757066/c 1628 bp DNA linear GSS 24-OCT-2003  
LOCUS P052-2-A02.za Ppa EcoRI BAC library Pristionchus pacificus genomic.  
DEFINITION genomic survey sequence.  
ACCESSION CG757066  
VERSION CG757066.1 GI:37985257  
KEYWORDS GSS.  
SOURCE Pristionchus pacificus  
ORGANISM Pristionchus pacificus  
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Nemodiplogasteridae; Pristionchus.  
1 (bases 1 to 1628)  
Srinivasan,J., Sinz,W., Jesse,T., Wieggers-Perebolte,L., Jansen,K.,  
Buntjer,J., van der Meulen,M. and Sommer,R.J.  
An integrated physical and genetic map of the nematode Pristionchus  
pacificus  
Mol. Genet. Genomics 269 (5), 715-722 (2003)  
12884007  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@uebingen.mpg.de  
Class: BAC ends.

FEATURES  
source  
1..1628  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"

ORIGIN  
/db\_xref="taxon:54126"  
/clone\_11b="Ppa EcoRI BAC library"  
/note="The library was generated by a partial digest of  
the genomic DNA with EcoRI and cloning into the BAC  
vector."

Query Match 1.34; Score 118.4; DB 10; Length 1628;  
Best Local Similarity 44.1%; Pred. No. 7.5e-11;  
Matches 639; Conservative 0; Mismatches 797; Indels 14; Gaps 5;

Qy 7543 CACCGAGCG 7602  
Db 1612 CCG 1553  
Qy 7603 GCCCG 7662  
Db 1552 NCCCG 1493  
Qy 7663 TTCTACCG 7717  
Db 1492 CCG 1433  
Qy 7718 AGCAGCG 7777  
Db 1432 CCG 1373  
Qy 7778 TCCCG 7837  
Db 1372 GCG 1313  
Qy 7838 CCGACCG 7897  
Db 1312 CCG 1253  
Qy 7898 CCGACCG 7957  
Db 1252 GCCCG 1193  
Qy 7958 CCG 8017  
Db 1192 CCG 1133  
Qy 8018 TGGCG 8077  
Db 1132 GCG 1073  
Qy 8078 ACCCG 8137  
Db 1072 CCG 1013  
Qy 8138 TCGAACCG 8197  
Db 1012 CG 953  
Qy 8198 GCGAGCG 8257  
Db 952 CCG 894  
Qy 8258 ACCG 8316  
Db 893 CCG 834  
Qy 8317 -CCCG 8375  
Db 833 CCG 774  
Qy 8376 CCG 8435  
Db 773 CCG 714  
Qy 8436 CCG 8495



Db	713	CCNCGCCGCCCCCGGCGCGGCCCCCCCCCGCCGCCCCCCCCCCCCCCCCCCCCCGGCG	654
Qy	8486	TTTTTTCATCGATGCGCTGTGAACAGCCATCTGTTCGGGCTGGTCCCGCTCGACAGCGCT	8555
Db	653	CCGCGCGCCCCCGCGCGCGCCCCCGCGCGCC-----CGCGCCCCCCCCCGGCGCGCGCGC	599
Qy	8556	TCGACGCGCGCGGACACGAGGTACGGGTCTCGGCGCTTCGCGGCGCTTGACCGACGTCAC	8615
Db	598	GCCCCCCCCCCCCCCGCCGC	539
Qy	8616	CCGATGCGCGGTCTGACCGC	8672
Db	538	GCCCCCCCCCCCCCCGC	479
Qy	8676	CCGACGCGCGCGCGGACGATCGTCGAGTACATGCGGACCGCTCGACTGAGTGCACAGAGC	8735
Db	478	CCCCCCCCCGGC	419
Qy	8736	ACACGACGATGTCCTGGGACCGACCTCTCGGGGATGACAGACCACTTCACCCCGACCTTCT	8795
Db	418	CGC	359
Qy	8796	TCGCGCTGATGAGCCCGGACCTCGCTCATGACAGCGGATGGTGAAGTT-CTGCGCGCTCGG	8855
Db	358	CGGC	299
Qy	8855	CGTCCGACCTGATGCTGTGGAGCGCGCTGACCTTCGCGCGCGCGCGCGCGCGCGCGCTC	8914
Db	298	CGCCCGC	239
Qy	8915	ACCGGACCCCGGACCGG	8974
Db	238	CGG	179
Qy	8975	AGCTTCTCGC 8984	
Db	178	CCCCCGCGCGC 169	
RESULT 6			
CC142184			
LOCUS			
DEFINITION	CC142184	987 bp	DNA linear GSS 16-Apr-2003
	NDL.80J12.T7	Notre Dame Liverpool	Aedes aegypti genomic clone
ACCESSION	CC142184		
VERSION	CC142184.1	GI:30011239	
KEYWORDS	GSS.		
SOURCE	Aedes aegypti	(yellow fever mosquito)	
ORGANISM	Aedes aegypti		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae; Culicinae; Culicini; Aedes; Stegomyia.		
REFERENCE	1 (bases 1 to 987)		
AUTHORS	Loftus, B., Shetty, J., Knudson, D. and Severson, D.		
TITLE	BAC end sequencing of Aedes aegypti		
JOURNAL	Unpublished (2003)		
COMMENT	Other GSSs: NDL.80J12.SP6		
	Contact: Brendan Loftus		
	Department of Bukaryotic Genomics		
	T1R		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-838-3543		
	Fax: 301-838-0208		
	Email: enthaefgr.org		
	Library was provided by David Severson		
	Seq primer: T7		
	Class: BAC ends.		
FEATURES			
source	location/Qualifiers		
	1..987		
	/organism="Aedes aegypti"		
	/mol_type="genomic DNA"		
	/strain="Liverpool"		
	/db_xref="taxon:7159"		

```

/clone="NDL.80312"
/clone.lib="Notre Dame Liverpool"
/notes=Vector: PECBAC1, Site1: Hind III, The library was
prepared from whole body tissue of newly hatched L1 larvae
by David Severson at the University of Notre Dame and
Hongbin Zhang"

```

Query Match 1.3%; Score 116; DB 9; Length 987;  
Best Local Similarity 47.8%; Pred. No. 2e-10;  
Matches 367; Conservative 0; Mismatches 400; Indels 1; Gaps 1;

QY	4135	CCGGGCGCAGGATGCTGCGAGCCGGGCA	CGGTGATGACCGGTGGCGGTGCGCCGACCA	4139
Db	202	CCGGGCTCTTAGTCGCTCTTAACAGTTCA	TCCCGGCTGACCGGATGAGGATCTTGAACA	261
QY	4195	GGATGATCACGTA	CTGGTAGTTGCTCTCTGTCGGGCGGAGCGAGTGCACGGTGCACG	4254
Db	262	GAATAGGGAATAATAGGCACTAGTTTGCAAC	CTTTTCCCGGCACTGACAGACACTGATATTC	321
QY	4255	CGGCTACGTGGCGAGCTCCGCTGGTGTAGAG	CCGGTGGTGAACCGGTTGTCTCCCTGG	4314
Db	322	CTGGAAACCTTCACAGAGCTGCAGATACAG	TTCATTAATTCCTGCTGTTTAGCAAA	381
QY	4315	TTCTCGGGAACGGTTCGAGGGAAGTGA	CCCCCATAGCCCGGCGGCACTCCGCTCATCTTGC	4374
Db	382	TACCATGATGATCGCTTTCAGCTGCAGAA	AGGCCACGCGGAGATTAATTCACCTCACTTATTC	441
QY	4375	CGTTGGTGCAGATCTCGGTGACCACTTGC	TCCGGGCGGATGCCGAGTTGTCATGAGCC	4434
Db	442	CATTGATTCCTGGGCGCTACCA	CAGTGTTCATTGACAAAGCCGAAATTTCTTAAGATGAT	501
QY	4435	GGATCCGTTGGGCGACAGGCGGTGTCGTG	CGGTGACACACGCCCCGCTCTCGAAGCGGTGA	4494
Db	502	CGATAGCGCTCTAATGTTTGGATCCGGG	CAGATGATGCTCGGCTTCGAAAGTGTGA	561
QY	4495	CCGCTTGTGGTGGGTGGAGACTGAACA	CTTCGCGCTGACCCGGATCCGACCAACCGACGTC	4554
Db	562	ATACCTTGTGTGATGGAGCTCAAC	CCGATTAATCAACATGGCGCAAAATCTTCC	621
QY	4555	CACCCGTCGTGCAAGCCAGGAGCGTGGG	CGGCGGTGCAAGAGACTTGAACCTGTGGTGG	4614
Db	622	CATCATACGCAACACGAAAGCAATGGCG	GCATCATATGATCACCCGTAAGTTGATGAT	681
QY	4615	CGGCGATCTTCGCGACGCTCTCCACA	AGTGGCTGTGCGGCCCAAGGTGATCCCGACGA	4674
Db	682	CTGCAATCTTTTGTATCTTATCTACAT	CGGAAGGTGTCCATTAACATGACAGGACATGA	741
QY	4675	TCGCGCCGGTCTCGCGGGTGCAGACG	CGGCGACGTGTTCGCGGTGCACCAACCGGTGC	4734
Db	742	TGGCTGTGCTGAGGAGGTGATGGCTGC	CTATCTGTCTCGGATTCATGTTCAGGGTTC	801
QY	4735	CCGGGTTCAGTGTGCGAACAACCGGTGT	AGATCCGAGCCACGCTGCGCGGTGG	4794
Db	802	GCGGATTCGATGACGAACACGCGGTTGA	TGTCATTCACACAAATGATATGCGAGTGG	861
QY	4795	CCGGAAGGTCAATTGA	CCGCATGATCACTTACCGGTGACGTCAACCGG	4854
Db	862	CACAAATGATATGGGTGTCTTATATG	CTTACACCGGTATTTCTCAATGCTTGAAGTGG	920
QY	4855	GTTTCACGGGCGACGTTGGCGTTTG	CAGATGGCGAATGCAATGCGCCGTACCC	4902
Db	921	TCAACAAGCCCATGTTTCATTGGAAA	ACAAGGCCAGGTGATTTACAC	968

SOURCE	Mus musculus molossinus (Japanese wild mouse)
ORGANISM	Mus musculus molossinus Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarchontia; Euarchontoglires; Glires; Rodentia; Sciurornathi; Muridae; Murinae; Mus.
REFERENCE	Abe,K., Noguchi,H., Tagawa,K., Yuzuriba,M., Toyoda,A., Kojima,T., Esawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaiki,K. and Shiroishi,T. Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis Genome Res. 14 (12), 2439-2447 (2004)  2 (bases 1 to 2332) 15574823
JOURNAL	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y. Direct Submission Submitted (17-Nov-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Suehiro-Chou,Tsurumi-Ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, url:http://hgsc.gsc.riken.go.jp/, Tel.:81-45-503-9111, Fax:81-45-503-9170) Clones are derived from the mouse BAC library MSGm01. For BAC library availability, please contact Kuniya Abe (abe@crc.riken.jp). Tsukuba Institute, Bio Resource Center, The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199 e-mail: abe@erc.riken.jp
COMMENT	PRIMERS Sequencing : TJ LIBRARY Vector : pBACE3_6 R.site 1 : EcoRI R.Site 2 location/Qualifiers  1..2332 /organism="Mus musculus molossinus" /mol_type="genomic DNA" /sub_species="molossinus" /db_xref="taxon:57486" /clone="MSGm01-15SA2.TU" /bex="male" /tissue_type="mixture of kidney and spleen" (clone_lib="MSGm01 Mouse Male BAC Library")
FEATURES	
source	
origin	
Query Match                 1.3%; Score 112; DB 10; Length 2332;	
Best Local Similarity       41.4%; Pred.No. 1.le=0;	
Matches   898; Conservative   0; Mismatches 1249; Indels   20; Gaps   7;	
OY	2424 CGTCGGCGCAGCCGCCGTTCGAATTGTTTGACCCGGGCAGCTCACCGACTTGCATCAAACTCGT 2483
Dd	2287 GTCCCCCGCGACGACGCGCCGCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2228
OY	2484 CGAGAGAACAATCTGTCCCCTCGCATCGGATCGAATCTTCGGCACCAGCTGACGAAGTCGT 2543
Dd	2227 GCACCCCGC 2168
OY	2544 GGACGACCGAATGACGGCTCCGGTTCGGGGTAGACCCGGCCGAGACGTAAGCGGTGGCC 2603
Dd	2167 GCCCCCG 2108
OY	2604 CCGCCAGCCCGGAAAACCGGCTCCGGTACAGGTAGACGTCCTCCAGCAGTCGATCT 2663
Dd	2107 CCGGCGCGCG-----CG 2055
OY	2664 GCACCGCACCCTGGAGATGGCGGTTCGGGCGCATGTTCCGGCCGGAATCCGCAACAGT 2723
Dd	2054 CGGCG 1995
OY	2724 GGGCGTCGACACCCGACGACGCTCTCAAAGCGATAACCAAGTCGATTCGATGCGCG 2783

[illegible]



Db 193 CCACCTGGGCGCAATGATATTGGCGGGTCAACGGGTCGATATCGCGCAACACCGGCGTGG 134  
QY 4765 GTCCGAGCCAGCTCGCGCGTGGCGCGGTGGCCCGGAGTCAATTGAGCGCATGACTT 4824  
Db 133 CGCGGACTTCCAGAAATACGTTGGCTGTGGCTACCCAGAAATCGGGGTGTGATGACTT 74  
QY 4825 CACCGGATGAGTCACCGGCGCGGAGCAGCAGTTCCAGGCGAGCGGTGGCGGTGAGGTGG 4884  
Db 73 CGTCCCAAGGCGCGGATACCGGCGATGGCGAAGGAGATCTCATGTGTGCAAGTGGCGGAGT 14  
QY 4885 CGATGCAGTG 4894  
Db 13 TGAAGGTGCG 4

RESULT 9  
AG381986 2243 bp DNA linear GSS 21-DEC-2004  
LOCUS AG381986  
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-192D21.TU, genomic survey  
sequence.  
AC381986  
VERSION AG381986.1 GI:47993191  
KEYWORDS GSS.  
SOURCE Mus musculus molossinus (Japanese wild mouse)  
ORGANISM Mus musculus molossinus  
Bukayota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS 1  
Abe,K., Noguchi,H., Tagawa,K., Yuzurika,M., Toyoda,A., Kojima,T.,  
Brawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaiki,K. and  
Shiroishi,T.  
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to  
genomic constitution of strain C57BL/6J, as defined by BAC-end  
sequence-SNP analysis  
JOURNAL Genome Res. 14 (12), 2439-2447 (2004)  
PUBMED 15574823  
AUTHORS 2 (bases 1 to 2243)  
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suehiro-cho,Tsukumi-Ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
COMMENT Libraries are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp

PRIMERS  
Sequencing : TU  
LIBRARY : PBACe3.6  
Vector : EcoRI  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.  
FEATURES  
source Location/Qualifiers  
1..2243  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-192D21.TU"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"

ORIGIN  
Query Match 1.2%; Score 110.4; DB 10; Length 2243;  
Best Local Similarity 42.5%; Pred.No.2e-09;  
Matches 514; Conservative 0; Mismatches 693; Indels 2; Gaps 2;

QY 7490 CCGCGGGA CCGGAGACCGTACGCCAGTCGAGCCGGAACGTGGGTGTGTGACACCGGA 7549  
Db 974 CCCCCCCCCCGGGGCGCNCNCGCGCNCGCGCCCCGCGCGCGCGCGCGCGCGCGCG 1093  
QY 7550 CCGCGCGCGAGGTGTCTGCGGATCCGAGCTTCAACCGAGCCCGCCGACGCTGCGGT 7609  
Db 1034 GCG 1093  
QY 7610 GATGTGAGGTGAGCCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7669  
Db 1094 GGGGGCG 1153  
QY 7670 CCGGACCGAGAGACGCGCGCGTCCGATGACAGTGAACGCGGACTGTTCACAGCGGTGCG 7729  
Db 1154 CCG 1213  
QY 7730 CAGAGCTGTGACCGAGCTGAGGTGCGCTTCGATTCGTGAACGACTTGGCCCGGAG 7789  
Db 1214 GCG 1273  
QY 7790 TCCCGGTGTGAGGCGCTGTGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7849  
Db 1274 CGCCCG 1333  
QY 7850 GATCTTGAACCTCGCGCGACCGCGGTATGCTTGAACGCGCGGTCAGCGCCGCAACGCTCG 7909  
Db 1334 CCG 1393  
QY 7910 CGGTGACCGAACAAGCGCTGACCGCCTCGACGAGATCGACGCGGTACCGCGGTGCG 7969  
Db 1394 CG 1453  
QY 7970 ACGCGCGGT 8029  
Db 1454 GGGCGGACG 1513  
QY 8030 TCTGTGCGCTGACCGAGACTTCCCGAATGCGCGGACGACTTGGCCGACCGGAGACCG 8089  
Db 1514 GCGGGGCG 1573  
QY 8090 CGACCGGT 8149  
Db 1574 GGGCGGGGCG 1633  
QY 8150 CCGCGCGGTGCG-AGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8208  
Db 1634 CG 1693  
QY 8209 GT 8268  
Db 1694 GCGGCG 1753  
QY 8269 GTGACCGGTGCG 8327  
Db 1754 GGGCGGACG 1813  
QY 8328 CTTGAGCGCCTGT 8387  
Db 1814 GGGGCG 1873  
QY 8388 GCGCGCGGTGCGCGGTGCG 8447  
Db 1874 CNGCG 1933  
QY 8448 TCTGAGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 8507  
Db 1934 GCG 1993  
QY 8508 TGGCTGTACAGCGCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8567  
Db 1994 CCG 2053

QY 8568 GACACGAGTACGGGTCTGTGCTCTCGCGGCGCTGACGACGAGCTGACCGGTGCTC 8627  
DB 2054 NCSCNCGNCGNCGNCGCGCCGCGCGCGGANNNGCGCGCGCGCCGCGCGGNGGAGNCGCGC 2113  
QY 8628 TAGCCCCCTGCCCCCTGCGTGTGAGAGCTGAACTTGTGTGAGTGTGACAGCCCGGAGCC 8687  
DB 2114 GCGCNAGGCGNCGCGGCG 2173  
QY 8688 AGGACATCG 8696  
DB 2174 CCGGCGCGCG 2182

RESULT 10  
AG363333  
LOCUS  
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-165A22.TJ, genomic survey  
AG363333 2332 bp DNA linear GSS 21-DEC-2004  
AG363333  
ACCESSION  
VERSION AG363333  
KEYWORDS  
SOURCE Mus musculus molossinus (Japanese wild mouse)  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCES  
AUTHORS  
TITLE  
JOURNAL

1 Abe, K., Noguchi, H., Tagawa, K., Yuzurika, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K. and Shiroishi, T.  
Contribution of Asian mouse subspecies Mus musculus molossinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis  
Genome Res. 14 (12), 2439-2447 (2004)  
1574823  
2 (bases 1 to 2332)  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
Direct Submission  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenho-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kunya Abe (abe@rc.riken.jp).  
Tsukuba Institute, Bio Resource Center,  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rc.riken.jp  
PRIMERS  
Sequencing : TJ  
LIBRARY : PBACe3.6  
Vector : EcorI  
R.Site 1 : EcorI  
R.Site 2 : EcorI  
Location/Qualifiers  
1.2332  
/organism="Mus musculus molossinus"  
/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-165A22.TJ"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC library"

ORIGIN  
Query Match 1.2%; Score 107.4; DB 10; Length 2332;  
Best Local Similarity 41.9%; Fred. No. 7.1e-09;  
Matches 709; Conservative 0; Mismatches 948; Indels 35; Gaps 5;  
QY 1194 CCTCCGACACCGCACCTTCTCTCCAGCCCAACCGCGTATCGAGGAGCGCGACCC 1253

DB 616 CCNCGNCGGACCCCACTTCTCCGCGGCGCGCTGCGCGCGGAGAGCGGCGCGCG 675  
QY 1254 GACCGCGGAGATATCAAGATTCAGACCCCGCGAGACCGGAGCCCTGCGCAAGTCT 1313  
DB 676 GCGTCCGCGTCCCGCGCAACCGCGGAGATCGCGCGCGCGCTTCCGCGCGCGCGCG 735  
QY 1314 CAGCAGCGCTTCAACCCCGGATAGATTCGCGCGAGCTTCAACCGCGC-----ATCGG 1365  
DB 736 GCGGCG 795  
QY 1366 GAGGTACCGCGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 1425  
DB 796 CCG 855  
QY 1426 GCTTTCGCGTCCGCGTCAAGATTCGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCG 1484  
DB 856 ACCCG 915  
QY 1485 CAGCAGTTCGCGTCAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1544  
DB 916 CCG 975  
QY 1545 TCCGCGCGTGTGCGAAGCGATCATGCGAGTGTGAACCGCGTCACTTCACTGTGCGA 1604  
DB 976 CCG 1035  
QY 1605 CAGGTTCGCGAAGCG 1664  
DB 1036 CCG 1095  
QY 1665 CAGGTTCGCGAAGCG 1724  
DB 1096 GCG 1155  
QY 1725 GCTCG 1784  
DB 1156 GCG 1215  
QY 1785 G-----CACTCGAGTACTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1827  
DB 1216 GCG 1275  
QY 1828 ATCGCGCGATCATGAGAGGTGTTGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1887  
DB 1276 CCG 1335  
QY 1888 ACCACGACGAGGCG 1947  
DB 1336 GCG 1395  
QY 1948 ACTCGGTCTCTCG 2007  
DB 1396 GCG 1455  
QY 2008 CCGTCCCGAAGATGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2067  
DB 1456 GCG 1515  
QY 2068 CTGCG 2127  
DB 1516 GCG 1575  
QY 2128 TAGGTGCGATGCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2187  
DB 1576 GCG 1633  
QY 2188 CTGCG 2247  
DB 1634 CCG 1693  
QY 2248 CCTTCGCGTCCGACCG 2307





OY		8256	CGACGGGTTTCAGCCGTGGAACCGTGGGCGGGAGAAGCGCAGATCTCTGTCTCCGGGGCGGCTC	8315
DB		1751	GCGGGGGCGGGGCGGGGCGGGGCGCCCGGCNMGCGGCCNAGCCCGGGGGGCGCGCG	1810
OY		8316	GCCCCGCAACGACTCGACGCGCTTGATGGCCACCTGTGACCAGCGGCGCGCTGCGGGCGCG	8375
DB		1811	CGAGNCGCCCG	1870
OY		8376	CGCGCGGGTGTTCGCCCGGCTGTCTCCGTTCCGGGCGCGGTGATCAGACGACGTGGTCACC	8435
DB		1871	CGCCCGCGANGGGCGGCGGCGCGCGCGGGGGGGGGCGCGCGGGGCGCGCCCGCGCGCN	1930
OY		8436	CGTGGCCC	8443
DB		1931	CGGCGCGC	1938
RESULT 12				
LOCUS	AG441877	1462 bp	DNA	linear GSS 22-DEC-2004
DEFINITION	Mus musculus mojosissinus DNA, clone:MSMg01-323H09.TJ, genomic survey			
ACCESSION	AG441877			
VERSION	AG441877.1	GI:48084940		
KEYWORDS	GSS.			
SOURCE	Mus musculus mojosissinus (Japanese wild mouse)			
ORGANISM	Mus musculus mojosissinus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Muridae; Euarctomoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 Abe,K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T., Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriaki,K. and Shirosishi,T.			
TITLE	Contribution of Asian mouse subspecies Mus musculus mojosissinus to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SNP analysis			
JOURNAL	Genome Res. 14 (12), 2439-2447 (2004)			
PUBMED	15574823			
AUTHORS	2 (bases 1 to 1462)			
TITLE	Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.			
JOURNAL	Direct Submission			
COMMENT	Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://ngp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@tc.riken.jp). The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan phone: 81-298-36-9189, fax: 81-298-36-9199 e-mail: abe@tc.riken.jp PRIMERS Sequencing : TJ LIBRARY Vector : PBACe3.6 R.Site 1 : ECORI R.Site 2 : ECORI. Location/Qualifiers 1..1462 /organism="Mus musculus mojosissinus" /mol_type="genomic DNA" /sub_species="mojosissinus" /db_xref="taxon:57486" /clone="MSMg01-323H09.TJ" /sex="male" /tissue type="mixture of kidney and spleen" /clone_tib="MSMg01 Mouse Male BAC Library"			
FEATURES	source			
ORIGIN				
Query Match	1.1%	Score 102.2;	DB 10;	Length 1462;

[illegible]





DEFINITION P052-2-A02.za Ppa EcoRI BAC Library *Pristionchus pacificus* genomic,  
genomic survey sequence.  
ACCESSION CG757066  
KEYWORD CG757066.1 GI:37985257  
SOURCE *Pristionchus pacificus*  
ORGANISM *Pristionchus pacificus*  
Bukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
Neodiplogasteridae; *Pristionchus*.  
1 (bases 1 to 1628)  
Srinivasan,J., Slinz,M., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,  
Bumfery,J., van der Meulen,M. and Sommer,R.J.  
An integrated physical and genetic map of the nematode *Pristionchus*  
*pacificus*  
Mol. Genet. Genomics 269 (5), 715-722 (2003)  
JOURNAL 12884007  
COMMENT Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
Classes: BAC ends.  
Location/Qualifiers  
1..1628  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Ppa EcoRI BAC Library"  
/note="The library was generated by a partial digest of  
the genomic DNA with EcoRI and cloning into the BAC  
vector."

ORIGIN

Query Match 1.1%; Score 101.4; DB 10; Length 1628;  
Best Local Similarity 44.7%; Pred. No. 8.5e-08;  
Matches 556; Conservative 0; Mismatches 669; Indels 20; Gaps 5;

QY 7448 GACGAGACGACGACCCCTTGAGACCTGAGTGAACGCTTTCGAGCGCGCGAGCGAGACCGT 7507  
DB 310 GCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 369  
QY 7508 AGCCAGTCCGCGCGGAGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7567  
DB 370 GGGAGGAGGAGCG 429  
QY 7568 CCGATCCGCGCTTCAACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7627  
DB 430 CCG 489  
QY 7628 GCGCGCGCGCGCTCTGCG 7687  
DB 490 CCG 549  
QY 7688 CGTCGTCGACGTCGACCG 7747  
DB 550 GGGAGGAGGAGCG 609  
QY 7748 TGGGAGTCCGCTTCACTGCGTGAACGACCTTCCCGCGCGCGCGCGCGCGCGCGCGCG 7807  
DB 610 CCG 669  
QY 7808 GTACCG 7867  
DB 670 GGGAGGAGGAGCG 7929  
QY 7868 CCGCGGTATGCTGACG--CCGAGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7925  
DB 730 GCGAGNAGGAGCG 789  
QY 7926 GCTGACCGCGCTCGACGAGATCGACGCGGCTCAACCGCGGTCGCGGACGCGCGGTCGTGCT 7985

DB 790 GCG 849  
QY 7986 GGGAGTGTGTCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8045  
DB 850 GCG 902  
QY 8046 GCTTCCCGAATGTCG 8105  
DB 903 GGGAGGAGGAGCG 962  
QY 8106 GAGAGTGTGTCGCGAGCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8165  
DB 963 CCG 1022  
QY 8166 CCGAGTGTGTCGCGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8225  
DB 1023 GGGAGGAGGAGCG 1082  
QY 8226 GAACTGTATCCCGAGCTTTCACGATCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8285  
DB 1083 CCGAGGAGGAGCG 1142  
QY 8286 CCGAGAGATCTGTGTCGTCG 8345  
DB 1143 GCGCG--CG 1196  
QY 8346 CACCTTGCGCGAGCG 8405  
DB 1197 CCG 1256  
QY 8406 CCG 8465  
DB 1257 GGGAGGAGGAGCG 1311  
QY 8466 CGAGCTGTAG 8525  
DB 1312 GCG 1371  
QY 8526 TGTTCGAGGTCGTCG 8585  
DB 1372 CCG 1431  
QY 8586 TCGCTCTGCG 8645  
DB 1432 GGGAGGAGGAGCG 1491  
QY 8646 GTACGACGTCGACCTTGTGAGTGCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8690  
DB 1492 CCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536

RESULT 15  
AG042920/c 1131 bp DNA linear GSS 01-NOV-2001  
LOCUS Pan troglodytes DNA, clone: PTB-021B06.R, genomic survey sequence.  
DEFINITION AG042920  
ACCESSION AG042920  
VERSION AG042920.1 GI:16571645  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes (chimpanzee)  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Pan.  
REFERENCE 1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
TITLE Direct Submission

**JOURNAL**

Submitted (02-ADG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenryo-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimpanzee@cc.riken.go.jp](mailto:chimpanzee@cc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>) Tel:81-45-503-9111, Fax:81-45-503-9170) Clones are derived from the chimpanzee BAC library PB3 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

**COMMENT**

Sequencing: M13Rev  
LIBRARY

Vector : PKS145

**R.Site 1 : Saci**

R.Site 2 : SACI.

## FEATURES

Location/Qualifiers  
1. .1131

1. 1131

```
/organism="Pan troglodytes"
```

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

```
/clone="PTB-021B06.R"
```

```
/bex="male"
```

```
cell_type="Lymphoblast"
```

```
/clone_lib="PTB Chimpanzee Male BAC Library"
```

**ORIGIN**

Query Match	1.14	Score 100.2	DB 10	Length 1131
-------------	------	-------------	-------	-------------

Best Local Similarity 45.0%; Pred. No. 1.4e-07;

Matches 406; Conservative 0; Mismatches 488; Indels 9; Gaps 4;

[illegible][illegible]

Search completed: April 7, 2006, 14:04:12  
Job time : 34756 secs



QY	4318	CGCGCAACGCGTCCAGGGAGGTGAGCCCAATGGCCCGCGGCGCACTGGCTATCTTGGCGGT	4377
Db	155672	CCTCCAGGCTCCGAAAGCTGGAAGATCCGATGACCGCGGCGAGTGTCCGTATCCGAAAGT	155613
QY	4378	TGTGTGCGGATCTCGGTGACCACTTGTCCGGGCGCAATGCCAAGTTGTCAATGGCCGGA	4437
Db	155612	TGTACCTTAGCTTCACGTGATGTGTAGCGTTCCGCGCTCGCGCTGTGTAGCGGAGATTCTGG	155533
QY	4438	TCCGTTCCGCGCAAGGCGCGTGTGTGTGACCAACGCGCCCGCGCTTCCAAAGCGGTGACCG	4497
Db	155552	CCCGGCTCGGCGAGTTCCCGCGTGTGTGTGTGTATCGCCCGCGCTCTCCCGTGTATGT	155493
QY	4498	CCTTGTGTGCGTGTGAAGCTGAACACTTCCGCGCTCACCGGATCCGCCCAACGACGTCCAC	4557
Db	155492	TCTTGTGTGGAATGAAGCTGAAGCAAGCGCGCGCTCTCCAAAGATCTTATCTCTTACCTT	155433
QY	4558	CCGTGTGCAAGCCCAAGGGCGTGTGGGCGCGGTGCAAGAAAGGCTTGAACCTGTGTGCGGCG	4617
Db	155432	TGTATTCGAGCTCTGTTGGGCGTGTGTGCGCGTCTTCCATCCAGATGATGATGTTCCTCCG	155373
QY	4618	CGATCTTGTGCGCAGCGCTCTCACAGGTGTGTGTGTGCGGCCCAAGTGTGACCCGACGATCG	4677
Db	155372	CGATCTTCAAGATTCGGGTTCATGTCAAGCGGGGTGTGCGTGAAGGTGCAAGAGGATCACTG	155313
QY	4678	CGCGCG-----TGTGCGGGGTGACGAGCGCGGCGAGCTGTGT	4713
Db	155312	CTACCGTGTGCGGGAGCTGAGCAAGCGCGGTCTTTGAGGAGCTTCAGACTCTCTCCAGGGAGC	155253
QY	4714	CGGGGTGACACGAGCCCGGTGTGCGGGGTGCAAGTGTGCAAGCACCGGTGTGAATCCGAGCC	4773
Db	155252	GTGATTCAGATACGTAAGTACGTTTCCAGATGATGTCCAGAAACGCAAGTGTGCGGCCACGT	155193
QY	4774	AGCTGTGCGCGCTGTGCGCGGTGTGCGCGCAAGGTCAATTGACGCGATCACTTCAACG----	4829
Db	155192	GTAAGCGCGCGTGTTCGAGTGAAGCATTAAGCTAAACCGCAAGATGCGAGGTCTGCGCG	155133
QY	4830	--GTACGTCAACCGGCGCGGACGACCAAGTTCCAGGGCGACGTTGGGCTTGCAGGTGGCGA	4887
Db	155132	GCCCCAGCCCGCGGAGCTCAAGGCGAGGTGCGGTGTGCGGCCGAGACTGTCTCGGA	155073
QY	4888	TGCAATGCGCGTACCCCGACCAAGTCCGGGACCCCGGCTCTGAACTCCCGACCAAGGGGCGC	4947
Db	155072	CGCAATGTTCAATCCGACGAAGAACTGAGGCGAACTCACGCTCGAACTCCTCAACCGTGTCTC	155013
QY	4948	C 4948	
Db	155012	C 155012	
RESULT 2			
US-11-143-980-1			
; Sequence 1, Application US/11143980			
; Publication No. US20050272133A1			
; GENERAL INFORMATION:			
; APPLICANT: He, Min			
; APPLICANT: Hucul, John			
; APPLICANT: Halcli, Bradley A.			
; APPLICANT: Wagenaar, Melissa M.			
; APPLICANT: Graziani, Edmund			
; APPLICANT: Summers, Mia			
; APPLICANT: Kulowski, Kerry			
; APPLICANT: Pong, Kevin			
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Production of a Complex			
; TITLE OF INVENTION: Polypeptide			
; FIDE REFERENCE: AM-101426US			
; CURRENT APPLICATION NUMBER: US/11/143,980			
; CURRENT FILING DATE: 2005-06-03			
; PRIOR APPLICATION NUMBER: US 60/664,483			
; PRIOR FILING DATE: 2005-03-23			
; PRIOR APPLICATION NUMBER: US 60/576,895			
; PRIOR FILING DATE: 2004-06-03			
; NUMBER OF SEQ ID NOS: 72			

```

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 116856
; TYPE: DNA
; ORGANISM: Streptomyces sp.
US-11-143-980-1

Query Match      1.4%; Score 122.8; DB 14; Length 116856;
Best Local Similarity 50.1%; Pred. No. 1e-15;
Matches 445; Conservative 0; Mismatches 402; Indels 39; Gaps 4;

QY 1259 CGGGAGTGAATCCAGAGATCGAACCCGCGCCGAGACCGGCGCCTGTGGCAAGGTCTGACGA 1318
Db 98628 CGGGCTTCTTCTTCCATGAGACCGCGCGACCAACACCGGCTAACCGCGCACACTCACCG 98687
QY 1319 GCGCCTTCAACCCCGGCTAGCATGCGGACCTTCGAACCGGCAATCCGGAGGTGACCCGGT 1378
Db 98688 CCGAGTCTTCGATGCGCGCGCGACGCGAATGACCGGCGGATTCGAGCGGCTGGCGGACC 98747
QY 1379 GCG-----TGTGTGGCGAAGCGCGGTGACCGGCTTGAACCTGTGTGAGGCGTCTGGCT 1429
Db 98748 GGGACCTCGATGCGATGAGAGCGGCGGCGACGAGCGCGGACCTCTGTGTGCGGCTAGCGCA 98807
QY 1430 TCCCGCTGACCGGTCAGCATCGTGTGCGCGAGCTGTGGGCGTGCCTCGCGATGACCAAGC 1489
Db 98808 GTCCGGTCCCGGATGTGTGATCTCCGAATCTCGGCGTGTGCTGTACACTACACAG 98867
QY 1490 AGTTGGTGACTGTGTCCGGCGGCTGTGTGACATTCAGATGAGCAACCCGACCGATCCGG 1549
Db 98868 AGTTGACGACGAGGTATCGCAGCTCCGGGAGACCGGCGGACGATCAGCGCTGTGCGCG 98927
QY 1550 CCTGTGTGAAAGCATATGACAGATGTGAACCCGCTACCTCTTACTGTCTGTGACAGAT 1609
Db 98928 CGATGCG-----GACCGCTGTGTGGAGCAGAGATCCCGGAAATTCG 98966
QY 1610 GTGGGGAACGGGCGGCGCGAACCCCGGAGACGACTGATCTCCGGGCTGTGTGCGCGAG 1669
Db 98967 TGTGTGCAACCGGCGCGAGCTCCGGGAGACGACTGATCAGCAGGCTGTGTGATGAG 99026
QY 1670 TCGACGGGCGCACCTTCGACGAGATGTGAGGCGGCGCACTTCTCCAGCGGTTGCTGCG 1729
Db 99027 TCGAGGGCGGTGTGCTGACCGACGAGAGAGTGTGTGCGCATTCGATCATTTTCG 99086
QY 1730 CGGGGCAATCAACCAACCGTCTGTGTGGCAATGTCTCGGACCTTCGACGAGACC 1789
Db 99087 CCGGTATGAAACCGTGTGAGAACCTTATCGGCTCTGGCATGTGTGCGCTGTTCCAGAG 99146
QY 1790 CGAGTATCGGACGGCGCGCGCGCGAGAACCGGCGTGTATTCGCGCGCATCGAGAG 1849
Db 99147 GTAGCGAGCTGACCCGGTGTGGGAGAACCCCGACTTATTCAGACGCGCGTGTGAGAT 99206
QY 1850 TGTTCG-----TTTCGCGCCCCCGCTTCCCCAGATGACGCGACACGACGAGCGCACCA 1906
Db 99207 TCTTTCGCTACTTCTCCCTGTAAACAATTGGGCAACGTTGGCAACCGCACCGAGATGCG 99266
QY 1907 CCGTGTGGGTGTGAGATCCGGGCGGACGTCAATGTCAACACTGTGGTCTCTGTGGCCA 1966
Db 99267 TGATCAATGTGACACCCATTCGAGAGGGGCGAGATCGTGTGCGGTGTGTCCACCGCCA 99326
QY 1967 ACCGCAATCCCCCTGTGGCGCATTCGGAACCGGACAGCTTTCGACCCGTCCCGCAAGATCGGT 2026
Db 99327 ACCCGGACCCCGAGCGTTCGCGGATCCGACCGCTTGTCTTCGACCGGTGACACA--- 99383
QY 2027 GTGCGCGGAGACTCTCTTGTGGGCAAGCGGTGACATTTGTGTGTGTCGCGTCCGCTGACGCG 2086
Db 99384 ---CTTCCACTCTGCGTTTGGGCAAGGTGTGACCAAGTGTCTGTGGGCGACGACTGGCGA 99440
QY 2087 GCTGTGAGAACGAGTGTGCTGTGAGAGATCATTCGCGCGGTAC 2130
Db 99441 GCGTGAACATGAGAGTGTCTCTACAGGCGTGTCTGTCAAGTTTC 99484

```





Db 5569 CTCGAGGCGGTGCGCTCAACACGCTGCGCGGAGTGTGTGATACCCGAGGTGCATCG 5628  
1485 CAGACAGTTGTGTGATCTGTGTCCGGCGCTGTGTGATCATCCAGATGAGACGCCGACCGA 1544  
Db 5629 CCGCGAGTTCTTGAAGATGTGTACCGATCTTGAACCTGCGCGCGCGCGGTGACCTC 5688  
Qy 1545 TCCGCGCTGTGTGAAACGATCAATGACAGTGTGAAACCGCTCACTTCACTGTCTGA 1604  
Db 5689 GAGAGAGACAGAGCGCGCGGTGTGAAGCAGAGCTGTGTGACATGTCTGACATCATGATGAG 5748  
Qy 1605 CAGGTGTGTGAAACG 1664  
Db 5749 GCTGTGTGAGAGAGCG 5805  
Qy 1665 CAGGTGTGAGAGAGCG 1724  
Db 5806 GAGCGAGAGAGAGAGCG 5865  
Qy 1725 GCTGTGTGAGAGAGCG 1784  
Db 5866 CCGCGCGCGAGATTCCTCACTTCCGTAACAGATGTGCGAACACCTGTCTCGCGCTCTCA 5925  
Qy 1785 GCACCGCGAGATCTGACG 1844  
Db 5926 CCAACCGCGAGAGCTCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5985  
Qy 1845 GAGAGTTGTGTGTTCG 1904  
Db 5986 GAGAGAGCTGT 6045  
Qy 1905 CACCGT 1964  
Db 6046 CAGATATCG 6105  
Qy 1965 CAACCGCGAGATTCCTGT 2024  
Db 6106 GAAACCGCGAGATTCCTGT 6165  
Qy 2025 TGT 2084  
Db 6166 -----CG 6219  
Qy 2085 GCAGCGT 2144  
Db 6220 CCGGATGAG-----AGTAAAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6270  
Qy 2145 CGACCGCGAGACGACGACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2204  
Db 6271 GGAAGCTGT 6330  
Qy 2205 CCGGTTGT 2232  
Db 6331 CTCGT 6358

RESULT 5  
US-10-497-135-17  
; Sequence 17, Application US/10497135  
; Publication No. US20050272132A1  
; GENERAL INFORMATION:  
; APPLICANT: Bioclica Technology Ltd  
; APPLICANT: Gregory, Matthew A  
; APPLICANT: Galsner, Sabine  
; APPLICANT: Petrovic, Hrvoje  
; APPLICANT: Mose, Steven  
; TITLE OF INVENTION: Production of Polyketides and Other Natural Products  
; FILE REFERENCE: 4408-P0344US00  
; CURRENT APPLICATION NUMBER: US/10/497,135  
; CURRENT FILING DATE: 2004-05-26  
; PRIOR APPLICATION NUMBER: PCT/GB03/003230  
; PRIOR FILING DATE: 2003-07-16  
; PRIOR APPLICATION NUMBER: GB0216509.0  
; PRIOR FILING DATE: 2002-07-16

; PRIOR APPLICATION NUMBER: GB0224922.5  
; PRIOR FILING DATE: 2002-10-25  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 17  
; LENGTH: 767  
; TYPE: DNA  
; ORGANISM: Streptomyces hygroscopicus  
US-10-497-135-17

Query Match 1.3%; Score 118; DB 8; Length 767;  
Best Local Similarity 50.1%; Pred. No. 1.3e-14; Indels 9; Gaps 2;  
Matches 351; Conservative 0; Mismatches 340

Qy 1459 CTGT 1518  
Db 6 CTGT 65  
Qy 1519 GACATTCAGATGTGACGACCCGACCGATCCGCGCTGTGTGTGTGTGTGTGTGTGTGT 1578  
Db 66 GAGGCGCAGATGTGCGGCGCATGTGAGGCGCGGACCGACATGTGAGATCAAGAGCTGACC 125  
Qy 1579 AACCGCTGT 1638  
Db 126 GAGGATTTCTGT 185  
Qy 1639 GACCTGT 1698  
Db 186 AGGCTTGT 245  
Qy 1699 GCGGCGCACTTCTGT 1758  
Db 246 GCAGCGGAGGT 305  
Qy 1759 GGCACATCTGT 1818  
Db 306 GGTACTGT 365  
Qy 1819 CCGGCTGT 1875  
Db 366 CCGGACCTGT 425  
Qy 1876 CAGATGT 1935  
Db 426 GCGGT 485  
Qy 1936 GT 1995  
Db 486 GACACGT 545  
Qy 1996 GACAGTTGT 2055  
Db 546 GACAGTTGT 599  
Qy 2056 GT 2115  
Db 600 ATCCACAAGT 659  
Qy 2116 ATCATGT 2155  
Db 660 CTGT 699

RESULT 6  
US-11-143-401-74  
; Sequence 74, Application US/11143401  
; Publication No. US2006002953A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of



```

/ TITLE OF INVENTION: the M. Tuberculosis Complex
/ FILE REFERENCE: STAN102CON
/ CURRENT APPLICATION NUMBER: US/11/143,401
/ CURRENT FILING DATE: 2005-06-01
/ PRIOR APPLICATION NUMBER: US/10/647,089
/ PRIOR FILING DATE: 2003-08-21
/ PRIOR APPLICATION NUMBER: US/09/894,844
/ PRIOR FILING DATE: 2001-06-27
/ PRIOR APPLICATION NUMBER: 09/318,191
/ PRIOR FILING DATE: 1999-05-25
/ PRIOR APPLICATION NUMBER: 60/097,936
/ PRIOR FILING DATE: 1998-08-25
/ NUMBER OF SEQ ID NOS: 137
/ SOFTWARE: FaestSeq for Windows Version 4.0
/ SEQ ID NO 74
/ LENGTH: 1215
/ TYPE: DNA
/ ORGANISM: Mycobacteria tuberculosis
US-11-143-401-74

```

Query Match 1.3%; Score 115.6; DB 14; Length 1215;

Best Local Similarity 49.5%; Pred. No. 4e-14; Mismatches 409; Indels 27; Gaps 4;

```

Matches 428; Conservative 0; Mismatches 409; Indels 27; Gaps 4;
QY 1279 GACCCGCGGAGACCGGCGCTGCGGAGGTGTCAGACGCGCTTCAACCCGCGTACG 1338
DB 277 GATCCGCGGTCACACCGAGTTTCGCACTGTGTGCGGCGGCTTCAACCGACGACG 336
QY 1339 ATGCGGACCTCGAACCGGCGATCCGAGAGTGAACCGGTGCTGTGCGCGA-----C 1392
DB 337 GTCGAAACCTTCAGCGCCACGCGCGCAAGTGTGTGTTGAGCGGCTCGAAAGCTGCGC 396
QY 1393 GCGGATGACCGCTTCAACCTGTGAGGCGCGCTTCCGCTCCGCTCCGCTCAAGATGTC 1452
DB 397 GCCAAGGTGCGCGCGACATTTGCAACGAACTATTCAACCGCTCCGCTGATGTGTG 456
QY 1453 GCGGAGCTGTGCGGCTGCGCGGAGTGAACCAAGACAGTTGCTGTGCTGCGCGCGC 1512
DB 457 GCGCATATCTCGGTGTGTCCCGAAGAGATTGACCGCAATTCACGCGGTGACCCAGGCC 516
QY 1513 CTGCTGACATCCAGATGAGACGACCGGATCCGCGCTGTGTGCAACGATCATGACG 1572
DB 517 ATGTGTGCGCGCAACCGGTTGACGCGCGCACACG-----GCGCATGAGAC 564
QY 1573 GTGCTGAACCGGCTCACTCTCTACCTGTGAGAGTGTGCGGAGCGGCGCGACCC 1632
DB 565 GCGGTGCGCTGATGATGCTTACTTCAACGCGCTGATGAGCGCGCGACCGACCC 624
QY 1633 CGGAGACACTGATCTCCGCGTGTGTGCTGCGCGAGT-----CGACGCGCGCACCC 1686
DB 625 GCGGACGACCGCATCTTCCACCTGTGAGCGCGCGGCGTCCGCGCGACCGACCGCC 684
QY 1687 GACGAGTGAAGCGGCGCACTTCTCAACGCTGTGCTGTGCGGCGGACATCACACC 1746
DB 685 GGCACACTGTCCATCTGCGGTTCACTTCAACATGATGACCGCGCGCAACGACCGTC 744
QY 1747 ACCGCTCTGTGCGGCAACATCTCTCGGACCTTGAAGAGACCGCGAGTACTGGAAGGCC 1806
DB 745 ACCGGAATGCTAGGCGGCTTGAAGCGGTGTGCAACCGGCGCGACCGACCGCGCTG 804
QY 1807 GCGCGGAGAGACCGGCGTGTGATCGCGCGATCATGAGAGAGTGTGCTTCCGCGCC 1866
DB 805 CTGCTGATGATCCCAAGAGGCACTCCCGACGCGGTGAGAGAGTGTGTCTGCGGCTCACCTG 864
QY 1867 CGGTTCCCGCAAGTGAAGCGGACACGACGAGCGGCGCAACCGTGTGTGTGAGTGAAGT 1926
DB 865 CCGGTGCAAGGCGGTGCGCGCAACGCGGAGTGAAGTGTGATGATGATGATGATGATG 924
QY 1927 CCGGCGGAGTGAAGTGAAGTGAAGTGTGCTGTGCGGCGCAACCGGATCCCTGCGGCGAT 1986
DB 925 CCGGCGGAGTGTGCGGCGGTGTGCTGTGCTGTGAGTGTGCGGCGCAACGCTGA---CGAACGCGCA 981
QY 1987 CCGGACCGGAGCACTTTCGACCGGTCCCGCAAGATGAGTGTGTGCGGCGAGCTCTCTTC 2046

```

```

DB 982 TACGCGCGGAGCGGACCGGACCTGATGATCTGCGGCGCGGCGCAACATCTTACCTTC 1041
QY 2047 GCGGAGCGGCTGACCTTGTCTGCGTCCCGCGTGTGCGGCGCTGAGAAACAGTGTGCC 2106
DB 1042 AGCCACGCGCGCGGACCGGACCTGCGGTGTGCGGCGCGCGCGGATGCAATCGGCTGTGCG 1101
QY 2107 CTGAGGAGATGATCGCGCGGTAC 2130
DB 1102 CTACCGAAGTGTGCGCGGTGTC 1125

```

# RESULT 7

```

US-10-497-135-18
/ Sequence 18, Application US/10497135
/ Publication No. US2005027132A1
/ GENERAL INFORMATION:
/ APPLICANT: Biotech Technology Ltd
/ APPLICANT: Gregory, Matthew A
/ APPLICANT: Galsner, Sabine
/ APPLICANT: Petrovic, Hrvoje
/ APPLICANT: Moss, Steven
/ TITLE OF INVENTION: Production of Polyketides and Other Natural Products
/ FILE REFERENCE: 4408-P0344US00
/ CURRENT APPLICATION NUMBER: US/10/497,135
/ PRIOR FILING DATE: 2004-05-26
/ PRIOR APPLICATION NUMBER: PCT/GB03/003230
/ PRIOR FILING DATE: 2003-07-16
/ PRIOR APPLICATION NUMBER: GB0216509.0
/ PRIOR FILING DATE: 2002-07-16
/ PRIOR APPLICATION NUMBER: GB0224922.5
/ PRIOR FILING DATE: 2002-10-25
/ NUMBER OF SEQ ID NOS: 63
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 18
/ LENGTH: 761
/ TYPE: DNA
/ ORGANISM: Streptomyces hygroscopicus
US-10-497-135-18

```

Query Match 1.3%; Score 112.2; DB 8; Length 761;

Best Local Similarity 51.8%; Pred. No. 2e-13; Mismatches 308; Conservative 0; Mismatches 278; Indels 9; Gaps 2;

```

QY 1564 ATCATGACAGTGTGTAACCGGCTCACTCTCTGACAGTGTGCGGAGCGCG 1623
DB 105 AGCATCAAGAGGCTGACCGACCAATTCGCTCACTCTGCGCTGTGTGCGGCGG 164
QY 1624 GCGGACCGCGGAGGAGCTGATCTCCGCGTGTGTGCGGAGTGTGAGCGGCGAC 1683
DB 165 GAGGAGCGGCGGAGGAGCTTCTGACCGCGGTGTGCGGAGTGTGAGCGGAGTGTG 224
QY 1684 CTGAGACGTAAGGCGGCGCACTTCTCAACGCTGTGCTGTGCGGCGGACATCAC 1743
DB 225 CTGAGAGGAGGAGGCGGCGCGGCGGTGTGCGGCTGTGTGTGCGGCGGCGGCGG 284
QY 1744 ACCACGCTCTGTGCGGCAACATCTCTGCGGACCTTGAAGAGACCGGAGTGAAG 1803
DB 285 ATGACAGCAATGTGTGTGCTGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 344
QY 1804 GCGGCGGCGGAGGAGCGGCGTGTGATGCGCGGATCATGAGAGAGTGTGTGCTTCCG 1863
DB 345 GCGCTGCGGAGAGACCGGAGCTGATGACGCGCGGTGTGAGAGAGTGTGTGCTTCTG 404
QY 1864 CCGCGTTCCCGGAG---ATGAGCGGCAACGAGCGGCGGCGGCGGCGGCGGCGG 1920
DB 405 CCGCTGACACAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 464
QY 1921 GAGATCTCGGCGGAGTGAAGTGAAGTGTGCTGTGCGGCGGCGGCGGCGGCGG 1980
DB 465 ACCATCAAGCGGCGGAGCAACGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 524
QY 1981 GCGCATCCGAGCGGCGGAGCACTTTCGACCGGTCCCGCAAGATGAGTGTGTGCGGCGGCG 2040

```



```
Db      890 AGCTCCCGCTCAAGTCCGCTTGCGCCCTTGAACGCGCGCAGCGCGGCGCGCTCC 831
Qy      2870 CGACCGCTGATCTCTTCCGCGAAGCTGACAAATCGAGCGCGCGCTGACTCG 2929
Db      830 TCCGCGCGCCGCGCAAGAGAGCTCTGCGCTGTCCTGGATGACAGCGCGCGCG 771
Qy      2930 TTGAGTTCCGCGATGAATGACCAAGGTCATGAGGCGGTTGCGCGCGCGCGAGATG 2989
Db      770 AGGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Qy      2990 ATGTGTTGAGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3049
Db      710 ATGCGCGAGGTCGCTGATTCACCAAGTTGCGCGCGCGCGCGCGCGCGCGCG 651
Qy      3050 CGGAGGTTCTTCTGACCCCGTTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 3109
Db      650 TGAACCCCGCTCGT-----GAAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Qy      3110 TCGTTGTTGAGCGCGCGTACAGGAGGTACGAGCGCGCGCGCGCGCGCGCGCG 3169
Db      599 AGGTACACCGCGCGCGCGCGCTGTTGTAAGCCGATGCGGAGCGCGCGCGCGCG 540
Qy      3170 CGGCGGAGCGTCTGTTCTGTTGAGGCGAGGCGCGCGCGCGCGCGCGCGCG 3229
Db      539 AGGATGAGCATAGCGAGGTTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Qy      3230 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3289
Db      479 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Qy      3290 CGGAGAGGTACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3349
Db      419 AGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy      3350 ATCTCGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3409
Db      359 TCGTACACGAGTACTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Qy      3410 AAGTGAAGCGGAAATGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3469
Db      299 AAGCGCGACAGCACTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Qy      3470 GGTCCCGCGGTACCGCAGCTCCAAAGCGCGCGCGCGCGCGCGCGCGCGCG 3529
Db      239 AAGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Qy      3530 CTCAGTACTTGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3584
Db      179 CGCCCTAGTCAGACGAGCGGCGCATGCGCGCGCGCGCGCGCGCGCGCGCG 125

RESULT 10
US-11-205-109-1/c
; Sequence 1, Application US/11205109
; Publication No. US20050287641A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Zazopoulos, Emmanuel
; APPLICANT: Staffa, Alfredo
; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
; FILE REFERENCE: 3002-2US
; CURRENT APPLICATION NUMBER: US/11/205,109
; PRIOR FILING DATE: 2005-08-17
; PRIOR APPLICATION NUMBER: US/09/976,059
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/239,924
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 88421
; TYPE: DNA

ORGANISM: Actinoplanes sp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (2077)..(3078)
OTHER INFORMATION: ORF 1; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (3118)..(4032)
OTHER INFORMATION: ORF 2; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (4038)..(5048)
OTHER INFORMATION: ORF 3; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (6665)..(5814)
OTHER INFORMATION: ORF 4; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (7703)..(6693)
OTHER INFORMATION: ORF 5; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (9464)..(8130)
OTHER INFORMATION: ORF 6; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (12751)..(10829)
OTHER INFORMATION: ORF 7; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 8; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 9; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (15591)..(15863)
OTHER INFORMATION: ORF 10; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 11; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 12; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (39713)..(65800)
OTHER INFORMATION: ORF 13; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 14; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 15; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 16; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (70059)..(70662)
OTHER INFORMATION: ORF 17; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (70059)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; negative strandedness
FEATURE:
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-11-205-109-1
```

Query Match 1.0%; Score 92.6; DB 14; Length 88421;  
Best Local Similarity 42.4%; Pred. No. 1.3e-09;  
Matches 1317; Conservative 0; Mismatches 1729; Indels 57; Gaps 13;

```
DB 2391 CGGCGGATCAGGCGCCGATCTGATCCACCCGTCGCGACGTCGCCGTTGAGTTGAT 2450
DB 47486 CGACGTAATCTCGGCGAGGCGCGGCGGCTCTTCAACCGGCGGAGATGAGCCGCG 47427
QY 2451 TGAGCGGCGGCTGACCGACTGTGTGAAACCGTCGAGGAGAACTGTCCCGGCTGCG 2510
DB 47426 TCTCTCGGCTTCAGCGGCTGACCTCGGCGGCGCGCGGATCGGCGGAGCA 47367
```

```
QY 2511 GATCGATGCTGCGGCCGACGTGACGAAAGTCGTGACGACCGAGTGCAGGCTCCG 2570
DB 47366 CGCGAGGAGACGCGGCTTACGTACCCGCGAGCGCGGCGCGTCCGCGCTCGAAGAGGT 47307
QY 2571 GGGTGAACCGCGCGCGGACGTAGCGGGTGGCCCCCGCAGCCCGGGGAAACCGGCTCCC 2630
DB 47306 CGGCGCGCGGAGTAGAGTTCACCGAGCGCGGCGCGGCTCCCGCGGCGCTCGAAGGTCT 47247
QY 2631 GTTACAGTGAAGTCTCCGAGGAGGTGATCTGACCGGCACTGCGAGTGGAGCGGTG 2690
DB 47246 CGGTGAACGTGAGGTGAGGTTCGAACTTGGCCGGAACCGGCGGCTGGACACGACA 47187
QY 2691 GGGCATCGTTTCGCGCGGATCCGAAACGCTGAGCGTGCACCCCGAACGAGCTCT 2750
DB 47186 TCGCCAGGCGCGGAGGTGCGGCGCGTCCGCGCGCGCGTCAAGGTGACATGACCT 47127
QY 2751 CCAACGCGTAACCGAGTGGGTGATGATGCGCCGCGGCTCGGCTGAGCGGCTGACGA 2810
DB 47126 GGAACAGGGGTGCGGCGGAGCGAGCGGCGCGGCTGAGCTCTTCGACCAAGCTTCTGA 47067
QY 2811 ACCGGGAGAGCCCTCCCGAATCGGTCCGTTCCGCTCGGACCACTGCGCGTCCC 2870
DB 47066 ACCGAGCTCTGAGTGGGCGAGCGCCGCAAGGTCAGCTGCGGACCTTGCGC-----A 47013
QY 2871 GACCGTGTGATCTCTCGCGGACGAGTGCAGAACTGAGCGGCGGAGTGGGCTGACTCGT 2930
DB 47012 GCACTGTGTCAGCGCGGCGGTTCGCGGTGAGTTCGAGCGGAGACGAGCGTGTGACGA 46953
QY 2931 TGAAGTTCGCGAGTGAAGTTCGACGAGTGCATGAGAGCGGTGGCCCCCGGCGGAGTGA 2990
DB 46952 AGAAGCGGACCAAGGTTCTCCAGCGCCCGGTGCGGCGGCGGCGGCGGCGGCGGCGG 46893
QY 2991 TGTGTTGAGGCGGAGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3050
DB 46892 GAGCTGTGAGTCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46833
QY 3051 GGAAGTCTTCTTGAACCCGTTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3110
DB 46832 TGAAGACGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46773
QY 3111 CGTTGTTCAAGCCGTCAGAGGAGTACGAGCCGCGGCGGAGGTCCACAGCGCGGCGTGC 3170
DB 46772 GGAATCTCGCGCTGTGACAGGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46713
QY 3171 GGCGCACGCTGTGTTCCGTGAGAGGCGGAGGCGGTTGTGTGAGACGTGAGGCGGAGCCCG 3230
DB 46712 CGGCGAGGTCAGGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46653
QY 3231 GCGGAGCGGCGTGCAGACCAACGCGCCGATACCCGAGTTGTGTCAGAGCGCGC 3290
DB 46652 GGGAGATCAGGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46593
QY 3291 CGAGAGGTACATCGCGGTGCGGTTTTCGCA--CGGCACTCTGTGATGACCGAGGTGAG 3348
DB 46592 AGTGGGTACTGTGACGCGGCGAGGCTCTCGGCTGCGGCGGCGGCGGCGGCGGCGG 46533
QY 3349 GATCTCGTTCGCGGCGGACCAACCGACTCGCGGTGTCATTGGCACTGTGACCGGTAGCA 3408
DB 46532 CGTGAAGCGGTGCGAGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46473
QY 3409 GAAATGACAGCGGAAACGACGTCGCG-----GCCGAGGTAGAGCCCGACGCTGACG 3461
DB 46472 GTTGCAGACACGACGACGAGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 46413
QY 3462 GAAAGCGGAGCTCCCGCGTACCGGAGCTTCGACGCGCGCGGCGGCTTACGCGGCTGGA 3521
DB 46412 GCGGAGATCTCCCGGCGGAGGTGGAAGGCTGACGCGCGGCTTCCGCGGCGGCGGCGGCG 46353
QY 3522 TGTGTTGTCATGATCTTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3581
DB 46352 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 46293
```

QY 3582 TGGGATCGAA CAGGTGAGCAGCTCGCCGAA CCGGCGCGGTTCAGTTCGAA CCGGCG 3641  
DB 46292 TGGTTCGCA CCGGCTCGCCGCTCGGCGA CCGGTAACA CCGTTCGAGCA CCGTTCGCGG 46233  
QY 3642 GCATTCCTCA CAGGCGGTGA CCGGCGAGCGCGTGA GTGGCGGTGA GTTTCGCGAGC 3701  
DB 46332 CCAACAGTGC CCGGAGCGCGCGGTG CAGGCGATCCCGGTGAGTTCCTCCGAGCGCA 46173  
QY 3702 GGAACCGCTGT CCGTTCGCGCGGTG CAGTTCGCGCA CCGGTGAGCGGTGAGC-----TC 3756  
DB 46172 GCGGACCGGAGATTTTGA GTGGCGCTCGGTCTTCAGCTGGGCA GAAACAGAGC 46113  
QY 3757 GGGCGTGA CCGCGCGCA CCA CCGCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGAT 3816  
DB 46112 GGGCGTTCGCGGTGA GAAAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 46053  
QY 3817 CTCCTGTGAT CCGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 3876  
DB 46052 GGTTCGTGCG CAGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 45993  
QY 3877 GTCGTGTGAT CCGGATCA GAAACCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCG 3932  
DB 45992 CCGGACCGAC CCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45933  
QY 3933 --ATCAACG CAGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCG 3990  
DB 45932 GCAAGCGATG CCG 45873  
QY 3991 CACTGAC CAGCGAGCG 4041  
DB 45872 GCACTTCG CAGAGCG 45813  
QY 4042 ACTTCGTGTG CAGCG 4101  
DB 45812 CACCGGCTTGA ACCGAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45753  
QY 4102 CCGGAGGAGAT GAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4161  
DB 45752 TCAAGCGG CAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45693  
QY 4162 CACGCTGAT GCGGTTCG 4221  
DB 45692 GCTTCG CAGTGGCG 45633  
QY 4222 CCTTCGT CCG 4281  
DB 45632 GCGGCTTGT CCG 45574  
QY 4282 AGAGCGGTGAT GAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4341  
DB 45573 --AACGCG CAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45516  
QY 4342 GCGCCATG CCG 4401  
DB 45515 CCTGCTGT CCG 44549  
QY 4402 TGTTCGCG CCGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4461  
DB 45458 GGTTCAG CCGGTTCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45399  
QY 4462 CCGTGA CCA CCG 4521  
DB 45398 CCGTGTG CCG 45339  
QY 4522 CCTTCG CCGGATTCG 4581  
DB 45338 CCGCGGTGA CCGG--ACCGG CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45282  
QY 4582 CCGCGGTGA GAGAGCTTGA CCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4641  
DB 45281 TGTTCG CAGCG 45222  
QY 4642 GTTCGTGT GCG 4701

DB 45221 TGGCGAGAG CGTTCACCTTCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45162  
QY 4702 CCGGACGTGT CCGGATTCGAGCA CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4761  
DB 45161 CGTTCAG CAGCG 45102  
QY 4762 TGAATCGAG CCG 4821  
DB 45101 TGAAGCGG CAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 45042  
QY 4822 CTTTCAC CCGGTGA CCG 4881  
DB 45041 GCAAGTGTG CACTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44982  
QY 4882 TGGCGATG CAGTCCG 4941  
DB 44981 CCGCGGTG CAGCA CAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44925  
QY 4942 GGGCGCG CCGGTTCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5001  
DB 44924 AGGAGCGT CAGAGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44865  
QY 5002 GCTG-----CCGTCG CCGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5055  
DB 44864 GGTTCGTGA CAGTTCGAGCA CCGTTCGATGAGTTCGATGAGTTCGATGAGTTCG 44805  
QY 5056 CCGGACCG CCGGAGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5114  
DB 44804 TGAAGCGG CAGGTTCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44745  
QY 5115 AGACCGG CAGCG 5174  
DB 44744 CAGCGCGG CAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44685  
QY 5175 CGAGCGAT TCGGATTCG 5231  
DB 44684 CCGGACCG CCGATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44625  
QY 5232 CGTGAGAT CCG 5291  
DB 44624 AGGAGAGT CCA CCGGAGAGAGCA CCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44565  
QY 5292 CACCTTC CCGAGTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5351  
DB 44564 GCACTGT CCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44505  
QY 5352 AGTCGAGT CAGTTCG 5411  
DB 44504 CCA CAGGCT CCG 44445  
QY 5412 GCA CCGTTC CCG 5454  
DB 44444 GGTTCG CCG 44402

RESULT 11  
US-11-075-185-2/c  
; Sequence 2, Application US/11075185  
; Publication No. US20050266434A1  
; GENERAL INFORMATION:  
; APPLICANT: REEVES, CHRISTOPHER D  
; APPLICANT: JULIEN, BRYAN  
; TITLE OF INVENTION: BIOSYNTHETIC GENES CLUSTER FOR AMERUTICINS  
; FILE REFERENCE: 010099\_03  
; CURRENT APPLICATION NUMBER: US/11/075\_185  
; PRIOR FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/551,103  
; PRIOR FILING DATE: 2004-03-08  
; PRIOR APPLICATION NUMBER: US 60/568,290  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 61





US-10-858-730-38

Query Match 1.0%; Score 89.2; DB 8; Length 2736;  
Best Local Similarity 41.8%; Pred. No. 8.2e-09;  
Matches 836; Conservative 0; Mismatches 1143; Indels 21; Gaps 4;

QY 3668 GAGGCGCGGTATGAGGCGGTGAGTTCCGCGAGCCGAGCCGCTGTCCGCGGTGCG 3727  
DB 2135 GAGGCGCGGTATGAGGCGGTGAGTTCCGCGAGCCGAGCCGCTGTCCGCGGTGCG 2076  
QY 3728 GATGCGGACCGGTGAGGCGGTGACCTTCCGCGTGAAGCCCGGACCAAGCGCGCG 3787  
DB 2075 GCGGCGGTGAGGAGGAGTCCATCCGCGCTCCAGCGGCGGACAGGCGCTGTCCGAGTGG 2016  
QY 3788 TCGGCGGTGAGGAGTCCATCCGCGCTCCAGCGGCGGACAGGCGCTGTCCGAGTGG 3847  
DB 2015 GCGGCGGTGAGGAGTCCATCCGCGCTCCAGCGGCGGACAGGCGCTGTCCGAGTGG 1956  
QY 3848 GTCCTTCAATGAGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCCGAGTCC 3907  
DB 1955 TCCGCGGCGAGGCGGCGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCCGAGTCC 1896  
QY 3908 TTTCCGCTTGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 3967  
DB 1895 TTGA---TCTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1840  
QY 3968 GTTCCGAGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4027  
DB 1839 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1780  
QY 4028 CCGTCCGAGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4087  
DB 1779 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1720  
QY 4088 CATCTGAGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4147  
DB 1719 GATACGCGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1660  
QY 4148 CCGCTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4207  
DB 1659 CAGCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1606  
QY 4208 CTGCTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4267  
DB 1605 GTGCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1546  
QY 4268 CAGCTGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4327  
DB 1545 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1486  
QY 4328 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4387  
DB 1485 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1426  
QY 4388 CTGCTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4447  
DB 1425 CAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1366  
QY 4448 CAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4507  
DB 1365 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1306  
QY 4508 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4567  
DB 1305 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1246  
QY 4568 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4627  
DB 1245 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1195  
QY 4628 CAGCGGCTTGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4687  
DB 1194 GAGCTGAGGCGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1135

QY 4688 CCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4747  
DB 1134 CAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1075  
QY 4748 GCAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4807  
DB 1074 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 1015  
QY 4808 TGAAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4867  
DB 1014 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 955  
QY 4868 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4927  
DB 954 GAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 895  
QY 4928 GAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 4987  
DB 894 GAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 835  
QY 4988 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5047  
DB 834 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 775  
QY 5048 GAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5107  
DB 774 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 717  
QY 5108 CCGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5167  
DB 716 GAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 657  
QY 5168 CTGCTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5227  
DB 656 AGGCTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 597  
QY 5228 CCGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5287  
DB 596 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 537  
QY 5288 CCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5347  
DB 536 AGGCTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 477  
QY 5348 AGGCTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5407  
DB 476 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 417  
QY 5408 CAGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5467  
DB 416 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 357  
QY 5468 AGGCTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5527  
DB 356 AGGCTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 297  
QY 5528 GCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5587  
DB 296 CCGTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 237  
QY 5588 AGGCTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5647  
DB 236 GTTCCGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 177  
QY 5648 CAGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 5667  
DB 176 TCGGCGGTGAGGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGCTTGTGAGTCCGAGTCC 157

RESULT 13  
US-11-205-109-1  
; Sequence 1, Application US/11205109



```
/ Publication No. US20050287641A1
/ GENERAL INFORMATION:
/ APPLICANT: Farnet, Chris
/ APPLICANT: Zazopoulos, Emmanuel
/ APPLICANT: Stafta, Alfredo
/ TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
/ FILE REFERENCE: 3002-2US
/ CURRENT APPLICATION NUMBER: US/11/205,109
/ PRIOR FILING DATE: 2005-08-17
/ PRIOR APPLICATION NUMBER: US/09/976,059
/ PRIOR FILING DATE: 2001-10-15
/ PRIOR APPLICATION NUMBER: US 60/239,924
/ PRIOR FILING DATE: 2000-10-13
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 1
/ LENGTH: 88421
/ TYPE: DNA
/ ORGANISM: Actinoplanes sp.
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (2077)..(3078)
/ OTHER INFORMATION: ORF 1; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (3118)..(4032)
/ OTHER INFORMATION: ORF 2; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (4038)..(5048)
/ OTHER INFORMATION: ORF 3; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (6655)..(5814)
/ OTHER INFORMATION: ORF 4; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (7703)..(6693)
/ OTHER INFORMATION: ORF 5; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (9464)..(8130)
/ OTHER INFORMATION: ORF 6; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (9691)..(10761)
/ OTHER INFORMATION: ORF 7; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (12751)..(10829)
/ OTHER INFORMATION: ORF 8; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (13617)..(12802)
/ OTHER INFORMATION: ORF 9; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (15203)..(13614)
/ OTHER INFORMATION: ORF 10; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (15591)..(15863)
/ OTHER INFORMATION: ORF 11; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (15880)..(19035)
/ OTHER INFORMATION: ORF 12; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (19032)..(39713)
/ OTHER INFORMATION: ORF 13; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (39713)..(65800)
/ OTHER INFORMATION: ORF 14; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65826)..(66530)
/ OTHER INFORMATION: ORF 15; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (66546)..(67370)
/ OTHER INFORMATION: ORF 16; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (67384)..(70059)
/ OTHER INFORMATION: ORF 17; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (70059)..(70662)
/ OTHER INFORMATION: ORF 18; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (70659)..(71906)
/ OTHER INFORMATION: ORF 19; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (73439)..(71964)
/ OTHER INFORMATION: ORF 20; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (74216)..(73563)
/ OTHER INFORMATION: ORF 21; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (75424)..(74213)
/ OTHER INFORMATION: ORF 22; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (75535)..(76464)
/ OTHER INFORMATION: ORF 23; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (78110)..(76449)
/ OTHER INFORMATION: ORF 24; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (79864)..(78107)
/ OTHER INFORMATION: ORF 25; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (81624)..(79861)
/ OTHER INFORMATION: ORF 26; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (81909)..(81682)
/ OTHER INFORMATION: ORF 27; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (82346)..(82062)
/ OTHER INFORMATION: ORF 28; negative strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (82587)..(84446)
/ OTHER INFORMATION: ORF 29; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (84481)..(85548)
/ OTHER INFORMATION: ORF 30; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (85556)..(86845)
/ OTHER INFORMATION: ORF 31; positive strandedness
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (87372)..(86803)
```

OTHER INFORMATION: ORF 32; negative strandedness  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (87454)..(88420)  
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only  
US-11-205-109-1

Query Match 1.0%; Score 87; DB 14; Length 88421;  
Best Local Similarity 44.3%; Pred. No. 1.8e-08;  
Matches 513; Conservative 0; Mismatches 625; Indels 21; Gaps 3;

```
QY 1083 CGAGAGGCGGTTCTGCTCACTGCGCGGTGATGCGCGACAGGACCCGGTCTGACA 1142
DB 61180 CGACGACCGTCTCAACCGCGGTGCGCGCGGTGATGCGCGACAGCTTACTGAGCGCGC 61239
QY 1143 GAGCGAGTACGGGCGTGGGACATCTTCCGACAGTGAAGTACGAGGCTTCCGCGA 1202
DB 61240 CAGCTGAGCCCGGCGCTATCTGAGCGCGCGCGCTGACCGGCGAGCGCTTCTGAGCGT 61299
QY 1203 CACCGCACCTTCTCTCCGACCCGCGCTCATATGAGGAGGCGGACCGGACGCGCGG 1262
DB 61300 CCGGTTGCGCCCGGCGGAGGCGGACATGACCGGACCGGCGGACCGGCTCGGAGAACGC 61359
QY 1263 GATGATTCAGAGATGACCGCGCGGACGCGGCGCGGCGGCGGCGGCGGCGGCGGCG 1322
DB 61360 CGACGCGGACCTGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 61419
QY 1323 CTTCACCGCGCGGATGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1382
DB 61420 TATCGAGCGCGGCGGAGGTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 61479
QY 1383 GCTGCGCGACGCGCGGTGACCGGCTTCTGACCTGATGAGGCGCGGCGGCGGCGGCGG 1442
DB 61480 GGTGCTGCGCGCGGAGGACATCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 61539
QY 1443 CAGGATGTCGCGGAGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1499
DB 61540 CGACCGGAGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61599
QY 1500 CTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1556
DB 61600 GCTGCGGCGGCTATCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61659
QY 1557 CGAAGCATCATGAGGTGCTGAACCGGCTCATCTGCTGACGAGGTGCTGCGGA 1616
DB 61660 CAAAGGAGAGCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61719
QY 1617 ACAGCGGCGGCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1676
DB 61720 CCGCGGCGGCGGCTACCGGCGGAGAGGTGCTGCGGAGGTCTTCCCGGAGGTGCTCGG 61779
QY 1677 GCGGACCTTCGACGAGGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1736
DB 61780 CCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61839
QY 1737 CATCACACACCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1796
DB 61840 CCGTCTGCTGAGTGAAGAGTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 61899
QY 1797 CTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1856
DB 61900 CCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 61959
QY 1857 TTTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1916
DB 61960 GCGCAACCTTCATCCCGGAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62019
QY 1917 GGTGAGATCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1961
DB 62020 GCTGAGAGGCGGAGATGACAGAGGTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 62079
QY 1962 GCGCAACCGGATCCCGTGGCGGATCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2021
```

```
DB 62080 GAGCGCATATATCCGCTCGCGCGGCTCCAGAGGCGATGCTTTCACGACCTATGAGC 62139
QY 2022 CGGTGATCGCGGCGGAGCTCTCTTCCGAGGACCGGCGGCGGCGGCGGCGGCGGCGGCG 2081
DB 62140 CGGCGAGACGCGGAGGACGCTTATCATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62199
QY 2082 GAGCGCGCTGAGAACGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2141
DB 62200 CCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 62259
QY 2142 CTGACACCGGAGGACGACGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2201
DB 62260 CACCGGAGGTGCTGAGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62319
QY 2202 CTTCCCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2220
DB 62320 CTTCCCGGCTGAGCGGAGGTG 62338
```

## RESULT 14

US-11-175-689-5  
Sequence 5, Application US/11175689  
Publication No. US2006024806A1  
GENERAL INFORMATION:  
APPLICANT: STINEAR, TIMOTHY P.  
APPLICANT: COLE, STEWART T.  
APPLICANT: LEADLAY, PETER F.  
APPLICANT: SMALL, PAMELA L.C.  
APPLICANT: JOHNSON, PAUL D.R.  
APPLICANT: JENKIN, GRANT A.  
APPLICANT: DAVIES, JOHN K.  
APPLICANT: HAYDOCK, STEPHEN P.  
TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING  
FILE OF INVENTION: NOVEL POLYKETIDES, THERAPY AND PROPHYLACTIC USES  
FILE REFERENCE: 03495.0329-01  
CURRENT APPLICATION NUMBER: US/11/175,689  
CURRENT FILING DATE: 2005-07-07  
PRIOR APPLICATION NUMBER: 10/987,592  
PRIOR FILING DATE: 2004-11-15  
PRIOR APPLICATION NUMBER: 60/519,864  
PRIOR FILING DATE: 2003-11-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn Ver. 3.3  
SEQ ID NO 5  
LENGTH: 1314  
TYPE: DNA  
ORGANISM: Mycobacterium ulcerans  
US-11-175-689-5

Query Match 1.0%; Score 86.6; DB 14; Length 1314;  
Best Local Similarity 46.7%; Pred. No. 2.9e-08;  
Matches 429; Conservative 0; Mismatches 469; Indels 21; Gaps 4;

```
QY 1252 CGAGCGCGGAGGATATCAAGAGATGACCGGCGGCGGACACCGGCGGCGGCGGCGGCGG 1311
DB 316 CTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 375
QY 1312 GTGAGGAGCGCTTCAACCGGCGGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1371
DB 376 GTGTCTCGGTGTTACCGGAAAGCAGTAGCGGATTAAGGAGATATGCAAGAGACT 435
QY 1372 ACCCGGTGCTGCTGCGG-----CGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1422
DB 436 GGTGCGGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 495
QY 1423 CTGCGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1482
DB 496 TACTGCTCCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 555
QY 1483 CAGAGCACTTCTGATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1542
DB 556 CCAAGACCGGT---TCTCAAGTTCTGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 612
```

QY 1543 GATCCGGCCCTGTGCGAACCATATCATGCTGAGTGTGTAACCCGGCTACCTCTCTAAGCTGTC 1602  
DB 613 ACATGGGTGACGACGACGAGGTGCGGACAGGGCTTCAAGAGACTTCACTTGTGATCAAC 672  
QY 1603 GACAGGTGTGCGGAACGCGCGGACCGACCCCGGAGCACTTATCTCCGGCTGTGCTG 1662  
DB 673 GAGCACTTCGAGAAATTCGGGTCTTACCCCGGTGACATCTGATGAGTCAAAATGATCAC 732  
QY 1663 GCCAAGTTCGACCGGCGCACCTCTGAGAGAGTGAAGGCGGCACTTCTCCACAGCGTGG 1722  
DB 733 GCAAGTAAATAATGCTCTCTCGAAAAACAACCTTCCACGGAACCGAATGCGGATGATGCGG 792  
QY 1723 CTGCTCCGCGGGGCAATC-----ACCAACACCGTCTCTGCGGCAACATCTGTCGAC 1776  
DB 793 CTGCTGTTGGCGGCGCACTTTCGCAACAACATGACCTGTTAGGCAACGGGATTCAGGTG 852  
QY 1777 CTCGACGACGACCGGAGTACTGACGCGCGCGCGGAGACCGCGGCTGTGATGCGCG 1836  
DB 853 TTGTTGACGCGCGCCGAACTGCGGAGCGCTTGAAGTACGCGCGCGCACTTTGCGCAAC 912  
QY 1837 ATCATCGAGAGGTGTGCTGCTTCCGCGCGCGCGCTTCCCGGATGCAAGCGACGACG 1896  
DB 913 GCGGTAGAAAGATCTGCGGTGTGAGCGCACCGGTTCACTCGCGGACGAAATGCGTCCG 972  
QY 1897 AGGCGACACACGCTCGGTGAGTGAATCCGCGCGGACGTCATGTCACACTTGGGTG 1956  
DB 973 AAGGACACCGAGGTGCGGAGTACCGCAATCAAGCGGCGCGAGCTGTGCGCATCTATCTG 1032  
QY 1957 CTCTGCGCAACCGCGATCCCTGCGCGCATCCGACCGGACACGTTGACCGCTCCG 2016  
DB 1033 GGGCGGTCACCGGTATCCGCGGTTCGCGGTTCGCGCATCCGATCCGCAACCGCTTGAATCA 1092  
QY 2017 AAGATCGGTGTGCGCGGCTCTCTTCCGCGCGCGCGGCACTTCTGTCTGCGTCC 2076  
DB 1093 GCCAAC---GCCAACCGGATCTGCAATCTCCGTTGCGCGGCGCGCACTTCTGCTGCGTCC 1149  
QY 2077 CCGGTGCGCGCTGTGAGAACGAGGTGCGCGCTGTGAGAGATCATGCGCGGTACGATCGA 2136  
DB 1150 GCCCTTGGCGCGGTGCGAAGGCGAATCGGATTTGAAATGCTTTCGACGCGCTTCCCTGAC 1209  
QY 2137 CTGCGCGTGCACCGCGACG 2155  
DB 1210 GTGCGCGCGCGGCGCGCG 1228

RESULT 15  
US-11-052-554A-529/c  
/ Sequence 529, Application US/1105254A  
/ Publication No. US2005028866A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Sachdeva, et al.  
/ TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
/ TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
/ FILE REFERENCE: 30853/40359A  
/ CURRENT APPLICATION NUMBER: US/11/052,554A  
/ PRIORITY FILING DATE: 2005-02-07  
/ PRIOR APPLICATION NUMBER: US 60/589,227  
/ PRIOR FILING DATE: 2004-07-20  
/ PRIOR APPLICATION NUMBER: IN 173/DBL/2004  
/ NUMBER OF SEQ ID NOS: 763  
/ SOFTWARE: PatentIn version 3.3  
/ SEQ ID NO 529  
/ LENGTH: 3240  
/ TYPE: DNA  
/ ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-529

Query Match 1.0%; Score 85.4; DB 14; Length 3240;  
Best Local Similarity 43.5%; Pred. No. 4,8e-08;  
Matches 629; Conservative 0; Mismatches 811; Indels 7; Gaps 5;

QY 7489 GCGCGCGGACCGGACCGTACCGCATGTGCGCGCGAAGTGGTGTGCTGATCAACACCGG 7548

DB 3030 GCCCGCGGTGTCACCTTTCGCGCGGTGTGTCGCGCGCGCGCACACCAACCGGCTCCCG 2971  
QY 7549 ACCCGCGCGAGGTGCTGCGCGATCCGCGCTTTCACCAACGCGCGCGCGCGCGCGCG 7608  
DB 2970 GGGCATCTCGGAGCTCGCGGTGCGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCG 2911  
QY 7609 TGGATGAGGTGCGCACTGCGCGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCG 7668  
DB 2910 ACCGAGCGCGCGGTGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2851  
QY 7669 GCCCGCACCGAGAGAGCGGTGTGCGGTGAGAGTGAAGCGCGCACTGAGCTCCAGAGCGGTG 7728  
DB 2850 GCGCGCACCGCGCGCTGCGCGCGGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 2791  
QY 7729 GCC-AGGCTGATGACCGAGCTGCGGTGCGCGCTTTCGATCTGTAACGACTTTCGCGCG 7787  
DB 2790 GCGAGGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 2731  
QY 7788 GATCCCGGTGCTGCGGTGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 7847  
DB 2730 GGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2671  
QY 7848 CCGGTCTTGAACCTGCGCGACCGCGGTATGCTGTAAGCGCGCGCGCGCGCGCGCG 7907  
DB 2670 GCCGACCTCGCGCGCTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 2611  
QY 7908 CGCGGTGACCGAAGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 7967  
DB 2610 AGTGCAGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2552  
QY 7968 GAGCGCGCGGTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 8027  
DB 2551 GCGCGGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2492  
QY 8028 GGTCTGCGGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 8087  
DB 2491 GGTGCGCGGTGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 2432  
QY 8088 CGCGACCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8147  
DB 2431 AGCGACCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2372  
QY 8148 CACCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 8207  
DB 2371 CACCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2315  
QY 8208 AGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 8267  
DB 2314 CGCGGTGACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2255  
QY 8268 GGTGACCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8327  
DB 2254 CCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2195  
QY 8328 CCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8386  
DB 2194 TCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2135  
QY 8387 TCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8445  
DB 2134 TCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2075  
QY 8446 GGTGACCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8505  
DB 2074 GGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 2015  
QY 8506 GATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8565  
DB 2014 CCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1955  
QY 8566 CGGACGAGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 8625

```

Db 1954 CGGCGCCACCGTTAGCACCAACACCGTTGCCGACGCGCGCGTGCATTGCGG 1895
QY 8626 TCTGACCGCGCGTGCCTCGGTGACGACGTGAACTTGTGAGTGGACCGCCCAAGGCGG 8685
Db 1894 AATTACGGTTGCCGCGCTTACCGGCATTGCGCACCTGTGCCGCGCGCGCTGCGCGC 1835
QY 8686 CCAGGACATGCTGATGATGACGAGACCTTGCATGAGGTGACGACGAGCCACACCAT 8745
Db 1834 CGCTGCCGCGCGTGCCTTCAACCTGCGCGCGCGCGCCACACGCGCAGCGTTACCGCTT 1775
QY 8746 GTCTGGGACGACCTCTGCGGATGACGACCACTTCAACCCCGACCTTCTTTCGCGCTGAT 8805
Db 1774 GTCCACCGGAGCGCGGAGACCGTGCACCGGACGATGCGGTGCGCGCGCGTGC 1715
QY 8806 GAGCCCGGACTCGCTCATCGACGGGATGTCGAGTTCTGCGGTCTGCGGTCCCGACT 8865
Db 1714 CGGACCGCGCGTGCCTCGCGCGCGCACCTGACCCCGGTCCCGCACCGGTGCGCGCTT 1655
QY 8866 GATCGTCTGGAGCGCGCTGACCTTTCGCGCGCGCGATCGCGGCGCGGCTCAACCGGACCCC 8925
Db 1654 GCGCGCGCGCGCGCACCTTTCGCGCGCGCGCGCGGTACCGGCGGCGCGCGGTGCGCG 1595
QY 8926 GCAAGCC 8932
Db 1594 CGGTGCC 1588

```

Search completed: April 7, 2006, 03:50:34  
 Job time : 3417.3 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: Apr11 7, 2006, 02:29:50 ; Search time 6263.93 Seconds  
(without alignments)  
11913.100 Million cell updates/sec

Title: US-10-611-442-1

Perfect score: 8942  
Sequence: 1 ggcgcgttcgacatccatccatg.....gtggagcaccggagatcc 9024

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 413468905 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8942	100.0	9024	US-10-611-442-1	Sequence 1, Appl1
2	1309.4	14.6	12441	US-09-861-289-3	Sequence 3, Appl1
3	1309.4	14.6	13613	US-09-861-289-3	Sequence 3, Appl1
4	1309.4	14.6	13613	US-09-860-846-3	Sequence 3, Appl1
5	1309.4	14.6	13613	US-09-836-821-3	Sequence 3, Appl1
6	1309.4	14.6	13613	US-10-271-889-46	Sequence 46, Appl1
7	1309.4	14.6	13613	US-10-398-605-3	Sequence 3, Appl1
8	1241.2	13.5	60136	US-10-205-032-1	Sequence 1, Appl1
9	1203.2	13.5	84428	US-10-229-148B-1	Sequence 1, Appl1
10	729.4	8.2	1565	US-09-793-708-24	Sequence 24, Appl1
11	729.4	8.2	1565	US-10-201-365-22	Sequence 22, Appl1
12	729.4	8.2	1565	US-10-160-539-24	Sequence 24, Appl1
13	729.4	8.2	1565	US-10-468-828-24	Sequence 24, Appl1
14	729.4	8.2	1565	US-10-468-828-24	Sequence 24, Appl1
15	722	8.1	1458	US-09-861-289-9	Sequence 9, Appl1
16	722	8.1	1458	US-09-860-846-9	Sequence 9, Appl1
17	722	8.1	1458	US-09-860-846-9	Sequence 9, Appl1
18	722	8.1	1458	US-09-860-846-9	Sequence 9, Appl1
19	722	8.1	1458	US-09-836-821-9	Sequence 9, Appl1
20	722	8.1	1458	US-10-398-605-52	Sequence 52, Appl1
21	720.6	8.1	1458	US-10-398-605-9	Sequence 9, Appl1
22	667.2	7.5	1467	US-10-647-196-1	Sequence 1, Appl1
23	612.8	6.9	17596	US-10-205-032-37	Sequence 37, Appl1
				US-10-611-442-2	Sequence 2, Appl1

C 24	596.6	6.7	1227	5	US-10-205-032-35	Sequence 35, Appl1
C 25	595.4	6.7	1248	3	US-09-861-289-7	Sequence 7, Appl1
C 26	595.4	6.7	1248	3	US-09-860-846-7	Sequence 7, Appl1
C 27	595.4	6.7	1248	3	US-09-860-846-7	Sequence 7, Appl1
C 28	595.4	6.7	1248	3	US-09-836-821-7	Sequence 7, Appl1
C 29	595.4	6.7	1248	3	US-10-271-889-50	Sequence 50, Appl1
C 30	595.4	6.7	1248	3	US-10-398-605-7	Sequence 7, Appl1
C 31	519.2	5.8	5970	3	US-09-793-708-21	Sequence 21, Appl1
C 32	519.2	5.8	5970	6	US-10-201-365-21	Sequence 11, Appl1
C 33	519.2	5.8	5970	6	US-10-160-539-21	Sequence 21, Appl1
C 34	519.2	5.8	5970	8	US-10-468-828-21	Sequence 21, Appl1
C 35	519.2	5.8	5970	8	US-10-468-828-21	Sequence 21, Appl1
C 36	500.4	5.6	45055	6	US-10-107-431-277	Sequence 27, Appl1
C 37	496.6	5.6	1401	6	US-10-107-431-40	Sequence 40, Appl1
C 38	478	5.3	5975	9	US-10-796-304-1	Sequence 1, Appl1
C 39	474.2	5.3	1404	6	US-10-084-846A-103	Sequence 103, App
C 40	474.2	5.3	59816	6	US-10-084-846A-1	Sequence 1, Appl1
C 41	474.2	5.3	59816	6	US-10-084-846A-2	Sequence 2, Appl1
C 42	463.4	5.2	30943	9	US-10-680-860A-1	Sequence 1, Appl1
C 43	461.2	5.2	1452	3	US-09-758-759-30	Sequence 30, Appl1
C 44	461.2	5.2	1452	10	US-11-021-825-30	Sequence 30, Appl1
C 45	461.2	5.2	1455	6	US-10-107-431-44	Sequence 44, Appl1

#### ALIGNMENTS

```
RESULT 1
US-10-611-442-1
; Sequence 1, Application US/10611442
; Publication No. US20040203015A1
; GENERAL INFORMATION:
; APPLICANT: Hutchinson, C. Richard
; APPLICANT: Katz, Leonard
; APPLICANT: Reid, Ralph
; APPLICANT: Hu, Zhihao
; APPLICANT: Gramajo, Hugo
; TITLE OF INVENTION: RECOMBINANT GENES FOR POLYPEPTIDE
; FILE REFERENCE: 300622009100
; CURRENT APPLICATION NUMBER: US/10/611,442
; CURRENT FILING DATE: 2003-06-30
; PRIOR APPLICATION NUMBER: US 60/393,016
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 9024
; TYPE: DNA
; ORGANISM: Micromonospora megalomicea
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Genomic DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(9021)
; OTHER INFORMATION: n = A,T,C, G, or none
US-10-611-442-1

Query Match 100.0%; Score 8942; DB 8; Length 9024;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 9024; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGTTCGATCAGCATGATCGCTTAATGTCGGTTCATTCCTTTGATGGGGAT 60
DB 1 GCGCGTTCGATCAGCATGATCGCTTAATGTCGGTTCATTCCTTTGATGGGGAT 60
QY 61 GATGATTCATTCGAAATTCGAGACCGGTGTAAGCTTCGGGTCGCTGTTTC 120
DB 61 GATGATTCATTCGAAATTCGAGACCGGTGTAAGCTTCGGGTCGCTGTTTC 120
QY 121 GTCAGTTCATTCGCGCGGAAAGACCGGAGCGGACGGAACGGAATGTTTCGGA 180
```

Db	121	GTCAATTCCCATGCGCGAAGAACCGCAACGAAACCGAAGAAATGTTCTCGCA	180
Oy	181	ATGTCCTCTGCTCAAGCGCGACCAAGAGTGGTTCGGGTCTCCGTCGTCTCCAGACTCTGG	240
Db	181	ATGTCCTCTCTGCTCAAGCGCGACCAAGAGTGGTTCGGGTCTCCGTCGTCTCCAGACTCTGG	240
Oy	241	GACGACGAGCCGCGGCTCAAGCAGGCTCAACCACTCTCAGACGTACATTCTCAATCTTCGCG	300
Db	241	GACGACGAGCCGCGGCTCAAGCAGGCTCAACCACTCTCAGACGTACATTCTCAATCTTCGCG	300
Oy	301	AAAGATGTTCCGTGGCGGCTCACCGGCTCTGCCCCGACGAGAGGTCAACCCGGAGTTCTGCTCATC	360
Db	301	AAAGATGTTCCGTGGCGGCTCACCGGCTCTGCCCCGACGAGAGGTCAACCCGGAGTTCTGCTCATC	360
Oy	361	ACCCGGGCGCGGCGGCTATCTGCTCCGCGGCTGACCGAGATCGCCCTCAAGCTCCGAGATAC	420
Db	361	ACCCGGGCGCGGCGGCTATCTGCTCCGCGGCTGACCGAGATCGCCCTCAAGCTCCGAGATAC	420
Oy	421	CAGCGTCTGATGTCGCGCGGCTGCGCGGCTCGGCTCGGCGACGACGTGACAGGCAACC	480
Db	421	CAGCGTCTGATGTCGCGCGGCTGCGCGGCTCGGCTCGGCGACGACGTGACAGGCAACC	480
Oy	481	CGCAGACTCAACCGAGGCGCTTCGGCTCTGGAGCGCGGCGCGGCTCTGTCGACAGTCCCGCTG	540
Db	481	CGCAGACTCAACCGAGGCGCTTCGGCTCTGGAGCGCGGCGCGGCTCTGTCGACAGTCCCGCTG	540
Oy	541	GGTGGGGGTCTGGAGTCCGAACGCTTCGCGAACTTGAGAGAGTCTTGGCTCATTGGCCAGAGAA	600
Db	541	GGTGGGGGTCTGGAGTCCGAACGCTTCGCGAACTTGAGAGAGTCTTGGCTCATTGGCCAGAGAA	600
Oy	601	TACCTGTCTGCGCGCGAAGTTGCGTCAAGGGATGTAACCGGAGGCGCTCATTCGACTTGACC	660
Db	601	TACCTGTCTGCGCGCGAAGTTGCGTCAAGGGATGTAACCGGAGGCGCTCATTCGACTTGACC	660
Oy	661	GCGCTCAACCGCGAGAGATTCGCTGCAAGAGGGGCTCCAGAGTGCAGTACATGCGGCGCTG	720
Db	661	GCGCTCAACCGCGAGAGATTCGCTGCAAGAGGGGCTCCAGAGTGCAGTACATGCGGCGCTG	720
Oy	721	CATCTCAGTGGTCCAGCGCGCGAGCGCTTGAGAGGCTTCCACCGGTTGCGGTCCGAACTC	780
Db	721	CATCTCAGTGGTCCAGCGCGCGAGCGCTTGAGAGGCTTCCACCGGTTGCGGTCCGAACTC	780
Oy	781	GTCCGACGAATCTGGCTCTGGAACCGGAGCGCGAGGTGCACACGATCCACACAGCGATCTG	840
Db	781	GTCCGACGAATCTGGCTCTGGAACCGGAGCGCGAGGTGCACACGATCCACACAGCGATCTG	840
Oy	841	AAACGCCGAGACCCGACTTCCAGGACGATCTGCGCGTCACTCCGTCCGTTCCGTCCGAGGTC	900
Db	841	AAACGCCGAGACCCGACTTCCAGGACGATCTGCGCGTCACTCCGTCCGTTCCGTCCGAGGTC	900
Oy	901	GCCGCGACAGATGGGGGTTCGGGTCGGGGTCCGGGCGAGCTGACCGAATTACCGGCTAGCGGC	960
Db	901	GCCGCGACAGATGGGGGTTCGGGTCGGGGTCCGGGCGAGCTGACCGAATTACCGGCTAGCGGC	960
Oy	961	GACACCTCTGAGCCCGCAATCAACGACATTTGGCGAAATCGACATCTGTGCCCGGGGGGGA	1020
Db	961	GACACCTCTGAGCCCGCAATCAACGACATTTGGCGAAATCGACATCTGTGTGCCCGGGGGGGA	1020
Oy	1021	CGGGTTGACGACGAACGATGGGGAGAACATGACCACTATCGAACAGATCCGACGATG	1080
Db	1021	CGGGTTGACGACGAACGATGGGGAGAACCAATGACCACTATCGAACAGATCCGACGATG	1080
Oy	1081	GCCGAGAGAGCCGTTCTGCTGACTGCGCTGGCGGGTGAAGCCGACAGGCAACCCGATCTGG	1140
Db	1081	GCCGAGAGAGCCGTTCTGCTGACTGCGCTGGCGGGTGAAGCCGACAGGCAACCCGATCTGG	1140
Oy	1141	CAGACACAGTACCGGCTCTGGGACATCTTCCGCGACAGTGAAGTACGGAGATCTCTCCG	1200
Db	1141	CAGACACAGTACCGGCTCTGGGACATCTTCCGCGACAGTGAAGTACGGAGATCTCTCCG	1200
Oy	1201	GAACACCGCACTTCTCTCCGACCCCAACCGCGTCACTGAAGGGGCGACCCGACGCGG	1260

[illegible]





Db	4501	TTGGTGGGTGGAAGCTGAACCTTCGGCGTCAACGGATCCGCCACGGACGTCACACCG	4560
Qy	4561	TCGTGACAGCCCAAGGCGGTGGCGCGGTTCGAAGAAAGACTTGAACCTGGTGGTCGGCGCGA	4620
Db	4561	TCGTGACAGCCCAAGGCGGTGGCGCGGTTCGAAGAAAGACTTGAACCTGGTGGTCGGCGCGA	4620
Qy	4621	TCCTTCGACAGCGCTTCACAGGTGCTGTGGCGCCCAAGGTGTAACGCCACGATCCGCGC	4680
Db	4621	TCCTTCGACAGCGCTTCACAGGTGCTGTGGCGCCCAAGGTGTAACGCCACGATCCGCGC	4680
Qy	4681	CGGTCTGCGGGGTGACAGAGCGCGGCAAGTGGTCGGGGTCGACCAAGACCGGTGCGGGT	4740
Db	4681	CGGTCTGCGGGGTGACAGAGCGCGGCAAGTGGTCGGGGTCGACCAAGACCGGTGCGGGT	4740
Qy	4741	CGAGTGGCGAAGAACCGGTGTGATGTCAGACCAAGCTTCGCGCGGTGGCGGTGGCGCGA	4800
Db	4741	CGAGTGGCGAAGAACCGGTGTGATGTCAGACCAAGCTTCGCGCGGTGGCGGTGGCGCGA	4800
Qy	4801	AGGTCAATTGACGGGATGATCACTTCAACCGGTGACGTCAACCGGCGCGACCAAGTTCGA	4860
Db	4801	AGGTCAATTGACGGGATGATCACTTCAACCGGTGACGTCAACCGGCGCGACCAAGTTCGA	4860
Qy	4861	GGGCGACGATGGCGGTGGAGGTGCGATGACAGTCCGATCCCGACCAAGTCCGCGACCC	4920
Db	4861	GGGCGACGATGGCGGTGGAGGTGCGATGACAGTCCGATCCCGACCAAGTCCGCGACCC	4920
Qy	4921	GGGCTTCGAATCTCCGACCAAGGGGCCCGGTGTGTGTAAGCAATGTTGTTTCAAGGCC	4980
Db	4921	GGGCTTCGAATCTCCGACCAAGGGGCCCGGTGTGTGTAAGCAATGTTGTTTCAAGGCC	4980
Qy	4981	ACTCCAGGCGGGGCGAGAACCGCTGCGCGGTGCGCATGTCGCGCGCGCCACGTCAGAGG	5040
Db	4981	ACTCCAGGCGGGGCGAGAACCGCTGCGCGGTGCGCATGTCGCGCGCGCCACGTCAGAGG	5040
Qy	5041	GGTGGAGGAACGCTTCGGGACCGCGGAAGATCGCCAGATCGATCGATCGGCTTCAAGC	5100
Db	5041	GGTGGAGGAACGCTTCGGGACCGCGGAAGATCGCCAGATCGATCGATCGGCTTCAAGC	5100
Qy	5101	CGTGGCGGTGTTGATGACGGCGGACCGCGAGGCTGCGCACTGATGTTGATGAC	5160
Db	5101	CGTGGCGGTGTTGATGACGGCGGACCGCGAGGCTGCGCACTGATGTTGATGAC	5160
Qy	5161	GTAAGTGGCTGGCGGAGCAGTTGGGTGATGCGCGCAAGGTGATCCACCGGTGGTCGGG	5220
Db	5161	GTAAGTGGCTGGCGGAGCAGTTGGGTGATGCGCGCAAGGTGATCCACCGGTGGTCGGG	5220
Qy	5221	GGCACCGGCGTCTGGAAGTCCGGGGGCACTCGACGAGCATGTACCGGTTCTCGTTGCG	5280
Db	5221	GGCACCGGCGTCTGGAAGTCCGGGGGCACTCGACGAGCATGTACCGGTTCTCGTTGCG	5280
Qy	5281	GTAAGACCGCCCACTCTCTCGAAGTCCAGCGCTGTGTAAGCGATCTGAGTACGGGAGC	5340
Db	5281	GTAAGACCGCCCACTCTCTCGAAGTCCAGCGCTGTGTAAGCGATCTGAGTACGGGAGC	5340
Qy	5341	GTCACGACGTAAGTGAAGGTGATGGGGGATGTTCTCGGGCAAGTGGGTGATGTTGTC	5400
Db	5341	GTCACGACGTAAGTGAAGGTGATGGGGGATGTTCTCGGGCAAGTGGGTGATGTTGTC	5400
Qy	5401	CGGCTGACAGTGAACCGTTCGCGCCCACTCGCGCAAGTGTGAGCCACCGCTTCGGTGGC	5460
Db	5401	CGGCTGACAGTGAACCGTTCGCGCCCACTCGCGCAAGTGTGAGCCACCGCTTCGGTGGC	5460
Qy	5461	CAGGTGACAGAGCGGTGAGGGTGCCTGATCTTCTTGAACAGAGAGGCGACGATGCG	5520
Db	5461	CAGGTGACAGAGCGGTGAGGGTGCCTGATCTTCTTGAACAGAGAGGCGACGATGCG	5520
Qy	5521	GTCGTGGCGGGGAGAGAGGTTGATCCAGGACCTCACTGGCGGTGACTGGT	5580
Db	5521	GTCGTGGCGGGGAGAGAGGTTGATCCAGGACCTCACTGGCGGTGACTGGT	5580
Qy	5581	CACGAGACCGCCGAAGATGCTGAAGTACTTCTGTCTCTGTTGATGTCGTCGTGCGG	5640
Db	5581	CACGAGACCGCCGAAGATGCTGAAGTACTTCTGTCTCTGTTGATGTCGTCGTGCGG	5640
Qy	5641	CCGAAATCAACCGCTGGGTGATGATGTCGCGCAACCGCGGGGTGTTGATGATGATGAC	5700
Db	5641	CCGAAATCAACCGCTGGGTGATGATGTCGCGCAACCGCGGGGTGTTGATGATGATGAC	5700
Qy	5701	TGTGTGCGTGAAGTCCGAGATCCAGCTCAGCAGGGTGTTCATGTTGTGACCGGGGTGCC	5760
Db	5701	TGTGTGCGTGAAGTCCGAGATCCAGCTCAGCAGGGTGTTCATGTTGTGACCGGGGTGCC	5760
Qy	5761	CGCCCCGACAGAAACCGACAGTGGGTCTGAAAGAACCCCTGGGGGAGTCCGTGACG	5820
Db	5761	CGCCCCGACAGAAACCGACAGTGGGTCTGAAAGAACCCCTGGGGGAGTCCGTGACG	5820
Qy	5821	TACCGGCGAGAGGCGCGGCTTCAACCTGTGTGATGATGACAGCCAGACCGACCGAG	5880
Db	5821	TACCGGCGAGAGGCGCGGCTTCAACCTGTGTGATGATGACAGCCAGACCGAG	5880
Qy	5881	GTCATGTTGACCAAGTGTGTCGTAAGGAGCATGCGCGACGCTGGGCGAGGTGACCA	5940
Db	5881	GTCATGTTGACCAAGTGTGTCGTAAGGAGCATGCGCGACGCTGGGCGAGGTGACCA	5940
Qy	5941	CTTGCAGTTGGGGTGTCTGGGGCGGTCTGAGAGACCTTGACCACTGTTGGCGTTGGC	6000
Db	5941	CTTGCAGTTGGGGTGTCTGGGGCGGTCTGAGAGACCTTGACCACTGTTGGCGTTGGC	6000
Qy	6001	TTTGCAGAGAACCAAGAACCTGCTCGAGCTGAGAGAGCTGACAGAGATCCCGGTGGG	6060
Db	6001	TTTGCAGAGAACCAAGAACCTGCTCGAGCTGAGAGAGCTGACAGAGATCCCGGTGGG	6060
Qy	6061	GGGGGTCCGTCGAAGTACTGATGAACTGGAACGGGATTCGTTGTGACACCTCATGTA	6120
Db	6061	GGGGGTCCGTCGAAGTACTGATGAACTGGAACGGGATTCGTTGTGACACCTCATGTA	6120
Qy	6121	GTTGCTGCGGGTGGCTCGAGGGGTGGCGGAGAGCTGGAACGGCGTTGATGTTGACAGGTTG	6180
Db	6121	GTTGCTGCGGGTGGCTCGAGGGGTGGCGGAGAGCTGGAACGGCGTTGATGTTGACAGGTTG	6180
Qy	6181	GGCTTTGGCTGACCAAGAGGTGACATGACCTGCGATCCCGGACGATCAACCGAG	6240
Db	6181	GGCTTTGGCTGACCAAGAGGTGACATGACCTGCGATCCCGGACGATCAACCGAG	6240
Qy	6241	GAACCCGATCTCGGGTGTGACGATGATGGTTGATCAAGTCCGTCGCTCCGAAAGTT	6300
Db	6241	GAACCCGATCTCGGGTGTGACGATGATGGTTGATCAAGTCCGTCGCTCCGAAAGTT	6300
Qy	6301	GGTCCGACGTCAGAGGCTCTGATGAGAGAAACCGCGCGGAGTCTGGCGGACGCGAC	6360
Db	6301	GGTCCGACGTCAGAGGCTCTGATGAGAGAAACCGCGCGGAGTCTGGCGGACGCGAC	6360
Qy	6361	GTCCTCGGGGTGAAACGACGACGCTTCAATGATGCTGTAAGGGCACTCGATGCACTCGAG	6420
Db	6361	GTCCTCGGGGTGAAACGACGACGCTTCAATGATGCTGTAAGGGCACTCGATGCACTCGAG	6420
Qy	6421	CCGATGCTCGGCGCGCGGTGGGCAACCAAGTCTGAGACGTCGTGGTGGTGGAGG	6480
Db	6421	CCGATGCTCGGCGCGCGGTGGGCAACCAAGTCTGAGACGTCGTGGTGGTGGAGG	6480
Qy	6481	TCGCGCGGTGCGAGTCAAGAAACGTAATGGTCCGATTTGTTGATCCGGAAGTCGATGAC	6540
Db	6481	TCGCGCGGTGCGAGTCAAGAAACGTAATGGTCCGATTTGTTGATCCGGAAGTCGATGAC	6540
Qy	6541	GTTGACCGGATCCCGCATACGCTCTCCGATGATGTCGTGGGCGGTCCGTGCGGTACCGC	6600
Db	6541	GTTGACCGGATCCCGCATACGCTCTCCGATGATGTCGTGGGCGGTCCGTGCGGTACCGC	6600
Qy	6601	CCGACGACCAATTCGTGATCAAGACCCCGCCAGTGTAGGAGCTTCGCTCCGACCGGAG	6660
Db	6601	CCGACGACCAATTCGTGATCAAGACCCCGCCAGTGTAGGAGCTTCGCTCCGACCGGAG	6660
Qy	6661	AAAGTCCGTGAAACAATTCGCGGTGACCGGTTCGCGGCTGCGGTGAAACCGGCGTCCGA	6720
Db	6661	AAAGTCCGTGAAACAATTCGCGGTGACCGGTTCGCGGCTGCGGTGAAACCGGCGTCCGA	6720

QY 6721 GCAACCCGATCATGCTGTCCGATGAACCTTCTAACTGTCCGCGCGCACATCTTTCTGACCG 6780  
 Db 6721 GCAACCCGATCATGCTGTCCGATGAACCTTCTAACTGTCCGCGCGCACATCTTTCTGACCG 6780  
 QY 6781 GTGTGTTCCGATGATGAACGCGTTCCGCGCGCGCTGTGAACTGTGCGTGGACCTGACCG 6840  
 Db 6781 GTGTGTTCCGATGATGAACGCGTTCCGCGCGCGCTGTGAACTGTGCGTGGACCTGACCG 6840  
 QY 6841 TTGCGCGGTGTTTCCGCGCTTCCGAACTGCGGAACTGCGTGAATGCGGCGAGGTGGAGCG 6900  
 Db 6841 TTGCGCGGTGTTTCCGCGCTTCCGAACTGCGGAACTGCGTGAATGCGGCGAGGTGGAGCG 6900  
 QY 6901 GGTGCTGACCGCGATGATCTGCAATCATGAGCGCTCAATGACATCTTTGTAGCATGCT 6960  
 Db 6901 GGTGCTGACCGCGATGATCTGCAATCATGAGCGCTCAATGACATCTTTGTAGCATGCT 6960  
 QY 6961 CCGCGCGGAGGTTCCGACAGGCGCGAAACCGCGCGCACTCGACCTGTTTCAACGCTGCA 7020  
 Db 6961 CCGCGCGGAGGTTCCGACAGGCGCGAAACCGCGCGCACTCGACCTGTTTCAACGCTGCA 7020  
 QY 7021 CATCACGATGCAAGCGCGATGACACCGACACCGCATGCTGTGCTGCGACCTGGAAG 7080  
 Db 7021 CATCACGATGCAAGCGCGATGACACCGACACCGCATGCTGTGCTGCGACCTGGAAG 7080  
 QY 7081 GTGCGCGATCAGGGAATGCGCGTGTCACTAGACAGCGCAACGCTGTCCGCGCT 7140  
 Db 7081 GTGCGCGATCAGGGAATGCGCGTGTCACTAGACAGCGCAACGCTGTCCGCGCT 7140  
 QY 7141 GCGGAAACAGCATCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7200  
 Db 7141 GCGGAAACAGCATCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7200  
 QY 7201 CCGTGTCAACGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7260  
 Db 7201 CCGTGTCAACGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7260  
 QY 7261 AGCGATGCAAGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7320  
 Db 7261 AGCGATGCAAGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7320  
 QY 7321 TCTTCAATCCGAGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7380  
 Db 7321 TCTTCAATCCGAGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7380  
 QY 7381 CAGATGCGCGCGGCTGTCTACTGCGGCGTGTGCTGCGACCGCGATCTGTACTGATGCTC 7440  
 Db 7381 CAGATGCGCGCGGCTGTCTACTGCGGCGTGTGCTGCGACCGCGATCTGTACTGATGCTC 7440  
 QY 7441 CTGTCCGGAACGGGACGACCGCTGGAACCTGGAACCTGGAACCTGGAACCTGGAACCTG 7500  
 Db 7441 CTGTCCGGAACGGGACGACCGCTGGAACCTGGAACCTGGAACCTGGAACCTGGAACCTG 7500  
 QY 7501 GGAACCGATGCGCGATGCGCGCGGACGCGGATGCGTGTGCGTGTGCGTGTGCGTGTGCG 7560  
 Db 7501 GGAACCGATGCGCGATGCGCGCGGACGCGGATGCGTGTGCGTGTGCGTGTGCGTGTGCG 7560  
 QY 7561 GTGCTGTGCGCGATCCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7620  
 Db 7561 GTGCTGTGCGCGATCCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7620  
 QY 7621 GCGCACTGCG 7680  
 Db 7621 GCGCACTGCG 7680  
 QY 7681 GACGCGCGCGTGTGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7740  
 Db 7681 GACGCGCGCGTGTGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7740  
 QY 7741 ACCGAGCTGGGCGTGTGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7800  
 Db 7741 ACCGAGCTGGGCGTGTGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7800

QY 7801 GCGCTGTGATCCG 7860  
 Db 7801 GCGCTGTGATCCG 7860  
 QY 7861 TCGGCGAACCGCGGATGCTGTGACGCTCCAGGTCAGCGCGCGCGCGCGCGCGCGCGCG 7920  
 Db 7861 TCGGCGAACCGCGGATGCTGTGACGCTCCAGGTCAGCGCGCGCGCGCGCGCGCGCGCG 7920  
 QY 7921 CAGCGCTGACCGCGCTTCCGACGAGTCAACCGCGTCAACCGCGTCAACCGCGTCAAC 7980  
 Db 7921 CAGCGCTGACCGCGCTTCCGACGAGTCAACCGCGTCAACCGCGTCAACCGCGTCAAC 7980  
 QY 7981 CTGTGCGGCGTGTGCGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8040  
 Db 7981 CTGTGCGGCGTGTGCGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8040  
 QY 8041 ACCGAGCTTCCGAACTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8100  
 Db 8041 ACCGAGCTTCCGAACTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8100  
 QY 8101 GTGACGAGGATGTGCGCGGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8160  
 Db 8101 GTGACGAGGATGTGCGCGGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8160  
 QY 8161 GACCGCGGCGTGTGCGCGGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8220  
 Db 8161 GACCGCGGCGTGTGCGCGGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8220  
 QY 8221 GCGGCGAACCGTGTATCCGAGGCTCTTCAACGATCCCGACCGCGTGTGCGAGTGTGCG 8280  
 Db 8221 GCGGCGAACCGTGTATCCGAGGCTCTTCAACGATCCCGACCGCGTGTGCGAGTGTGCG 8280  
 QY 8281 GCGGCGAACCGTGTATCCGAGGCTCTTCAACGATCCCGACCGCGTGTGCGAGTGTGCG 8340  
 Db 8281 GCGGCGAACCGTGTATCCGAGGCTCTTCAACGATCCCGACCGCGTGTGCGAGTGTGCG 8340  
 QY 8341 GTGCGCAACCTGCG 8400  
 Db 8341 GTGCGCAACCTGCG 8400  
 QY 8401 CGTTCG 8460  
 Db 8401 CGTTCG 8460  
 QY 8461 CCGGTGAGCTGTGAGGAGAAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8520  
 Db 8461 CCGGTGAGCTGTGAGGAGAAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8520  
 QY 8521 CCACTGCTGCG 8580  
 Db 8521 CCACTGCTGCG 8580  
 QY 8581 GGTGTGTGCG 8640  
 Db 8581 GGTGTGTGCG 8640  
 QY 8641 CCGTGTGAGCGAGTGTGAGAACTTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCG 8700  
 Db 8641 CCGTGTGAGCGAGTGTGAGAACTTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCG 8700  
 QY 8701 GTTACATGCGGAGCCCTGCACTGCGGTGCAACGAGCGCAACACATGCTCTGCGGACGAC 8760  
 Db 8701 GTTACATGCGGAGCCCTGCACTGCGGTGCAACGAGCGCAACACATGCTCTGCGGACGAC 8760  
 QY 8761 CCGTGTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 8820  
 Db 8761 CCGTGTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 8820  
 QY 8821 CATCGACGAGATGTGTGAGTGTGCGCGCTCTGCGCGTCTGCGAGCTGTGCTGCGAGCG 8880  
 Db 8821 CATCGACGAGATGTGTGAGTGTGCGCGCTCTGCGCGTCTGCGAGCTGTGCTGCGAGCG 8880  
 QY 8881 GGTGACCTTTCG 8940

|||||  
Db 8881 GGTGACCTTGGCCGCCCCGATTCGGGCGCCGGGTCAACGGAAACCCGACGCCGGATGCT 8940  
Qy 8941 GTGGGGTTCGGGACGTCCGACACCCGGGGCCCGGACAAAGTTCTCTGGATCTGTGGCCCA 9000  
Db 8941 GTGGGGTTCGGGACGTCCGACACCCGGGGCCCGGACAAAGTTCTCTGGATCTGTGGCCCA 9000  
Qy 9001 GGAAGTGAAGACCGGGAGATCC 9024  
Db 9001 GGAAGTGAAGACCGGGAGATCC 9024  
  
RESULT 2  
US-09-988-384B-3  
Sequence 3, Application US/09988384B  
Publication No. US2003007824A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.536US1  
CURRENT APPLICATION NUMBER: US/09/988.384B  
CURRENT FILING DATE: 2001-11-19  
PRIOR APPLICATION NUMBER: PCT/US99/14398  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: US 09/105,537  
PRIOR FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 53  
SEQ ID NO 3  
LENGTH: 12441  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-988-384B-3  
  
Query Match 14.6%; Score 1309.4; DB 3; Length 12441;  
Best Local Similarity 67.7%; Pred. No. 8.4e-284;  
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;  
  
Qy 2369 GGGCGGATCGACGCGCTTCAGACGGCGGATCAAGCGCCCGATGCTGATCCACCGCTCG 2428  
Db 9472 GGGGGGATCGGGGCGGGGTGCGGGTCAAGCGCAAGAACCGCGGCGCTCTCCCAACCGTCC 9531  
Qy 2429 GCGACGTCCCGTTCGAGTTGTTGAGCGGGCGGTCACCGACTGTGCGAAACCGTCAAG 2488  
Db 9532 GCGGCGTCCGCTCCAGCTGTTCAGCGGGCGGGTGAACCTTATTCAGAACCGCTCATG 9591  
Qy 2489 AAGAACTGTCCTCCCGCGCTGCGGATCGATGCTGCGCGCCGACGTGACGAAAGTCTGAGCG 2548  
Db 9592 AAGTATCTGTGCGCGTTCGACGGCGGACACTGCGCGCGCGCTGACGAAATCTCTGACG 9651  
Qy 2549 ACCGATGCGAGCTCCGCTCCGGTCCGGGTGAACCGGCGCGGACGTAGCGGGTGGCCCGCC 2608  
Db 9652 ACTCGGTGAGGAGGTGTCCGGGGGTGACGCGGCGCGGATGTAGCGGGTCCGCGCGTCC 9711  
Qy 2609 AGCCCGGGGAAACCGGCTCTCCCGGTACAGGTAGAGTCTCCGAGGAGGTCCATCTGACCC 2668  
Db 9712 AGGTGCGGGGAAACCGGCTCTCCCGGTACAGGTACAGTTCGCGAGGAGATTCGACTGACC 9771  
Qy 2669 GCGACCTGCGATGGGCGGTGCGGGCGCATGCTTCCGGCGCGGATCCGAAACGACTGCGCG 2728  
Db 9772 GCGACCTGCGGTGCGCGGTGCGGGCGCATGCTTGGCGGGCTTGATTCGAGACGATTCCGCG 9831  
Qy 2729 TCGAACCCCGACGCGAGCTCTCAACCGGTAAACCAAGTGGTGTGATGCGATGCGCGGGTCC 2788  
Db 9832 TCGGCCCCGGGTGCGGAGGCTGTTCAGGGGTAGCGGTGATGATGAGTCCGGGGGTG 9891  
Qy 2789 CGCTGCGGGCGGTGATCGACGAACCGGGCGAGGCGCTCCGGAATCTCGGTCTGCTCCGCC 2848  
Db 9892 CGCTGCGGAGACCGGCTCTCTCGAAGGCGTTGAGGGCTCTCTGAGGTTCGGCCGCGCTCTCC 9951  
Qy 2849 TCGACGAACCTGCGGTCTGCGGACCGGCTGTAGTCTCTGCGGACGGTGACGAAGTCAAGC 2908

|||||  
Db 9952 TCGGCAAGCTTGCCTGCTGTCACGGCGCGCTGTAGTCTCCGGAATGTTGAGGAATGCATC 10011  
Qy 2909 GGCCTGTGCGGGCTGATCTGCTTGAAGTTCGGCGATGAAGGTGACAGAGTGAAGAGCGG 2968  
Db 10012 GTCTTGTCTGCTCCCGGCGGTGTGAGAGTCCGCAATGAAGTGAACAGAGTGAAGAGCGG 10071  
Qy 2969 TTGGCCGCGCCCGGACAGATGATGTGTTGAGGCGGAGCGGACCGGTGCGCGCTTCCG 3028  
Db 10072 GAGGCAACGCGCCGAGGACAGATGTAGCGGAACGCGAGTTGATTCGCGCATCTGCGCTCG 10131  
Qy 3029 GCGCGCATCTCGCAGAAAGTCCCGAGGTTCTTCTGACCCGTTTGAAGGCGCCACGTTG 3088  
Db 10132 GCGCGCAGCTGTGAGAAAGCGCGCGAGGTTCTTCGCGACGCGGCGGAGAGCGGCTTCTTG 10191  
Qy 3089 CCGGTGTGCGCTGTGATCTGCTGTGTTGAGCGCGGACAGGAGGATGAGGACCGCGCGC 3148  
Db 10192 CCGGTGTGCTGTGATCTCTGCTGTGAGGCGGTGAGACGAGGTGCGGATGAGCGTGC 10251  
Qy 3149 AGGTCCCAAGGCGCGGCTGCGCGGCGAGCGTCTGTTGAGTGAAGGCGAAAGCGTTGGT 3208  
Db 10252 AGGCCCCAGAGGCGCGGCTGCGCGTCCAGGGTGTGCTCGGTGAGCGGAGGAGTTGCTG 10311  
Qy 3209 TAGACGCTCAGGGCGAAAGCCCCCGCGCGCGGTGCGACACCAAGCCCGATACCGCGG 3268  
Db 10312 TAGACGCTGAGCGCGCAGAGCGGTGCTGAGTGGCGCGCGCAGGCTCCCGAGGCGCGGG 10371  
Qy 3269 TTGGTCAACGCTTCCAGGCGCGCGGAGAGGATCAATGCGCGTGGGTTCTCGAACGCGAC 3328  
Db 10372 TTGGTGAAGGAGCTCAAGCGCGCGGAGAGATCAATGCGCGAGGAGTTCGCGGAGTATC 10431  
Qy 3329 TCGTGTATGACCGGAGTGAAGATCTCGTTGCGCGGACCAACCGACTCGCGTCTGACTTG 3388  
Db 10432 TCGTGTATGACCGGACCGGACATAGCGTTCCGCGGTGAGAGGCGAGCGGATCTGACCG 10491  
Qy 3389 GCACTGTTCACCGGTACGCAAGAGTGAACAGCGGAACTGCACTGCGGCGCGGTTAAGC 3448  
Db 10492 GCGCGGATCAACGAGACGAGAAAGTGCACGCGGAACTGCAAGGTTCGCGCGCGGTAAAG 10551  
Qy 3449 CCGACGCTGTAACGGGAAAGGCGGCTCCGCGCTGCGCGACGCTCAACGCGCGCGGCG 3508  
Db 10552 CCGACGCTGTAACGGGAAAGAGCGGCTTCTGCGAGGCGCGCGGTGAGAGACGCGCGCTGT 10611  
Qy 3509 TTCAAGCGGCTGATCTGTTGCTCCAGTAATTCGCGCGCGGCGCTCTCACCGCGTGG 3568  
Db 10612 TCAGCGGAGAGAGGAGGTGTTCTTCAGTACGCCCGCGGCGCGGTCTGACCGCGGTG 10671  
Qy 3569 CCGAGCTCGGAGATGCAATGCAACAGGTTCAGACGCTCGCGAAACCGGCGCGGTCCAGG 3628  
Db 10672 CCGAGCTCCGGGACCTGCGCGAACAGGCGAGGCGCGCGGAGAGCGTCCCGGTGACG 10731  
Qy 3629 TCGAACCGAGCGGCGATCTGCTCCAGCGGGTGAACGCGGAGGCGCGGTAGTGGCGGTG 3688  
Db 10732 CCGAGGTGTAAGCGGCGCTCTCCAGCGGGTGAAGGCGCTGTTGCGGTACGACGCG 10791  
Qy 3689 AGTTGCGAGGCGGAGCGGCGCTGTGTTGCGGCGGTTCGAGTGCAGCACCGGTAGAGCGG 3748  
Db 10792 AGCCGAGAGAGTGAAGCGGCGGTGCTTCGCGCTGTGCGGCGGACGAGGCGCGCGCG 10851  
Qy 3749 GTGACTTCGCGCTGAGAGCGCGGACCAAGCGCGCGCGGTGCGGCTGAGTCTGTCGCG 3808  
Db 10852 GCGAGGAGTGAAGCGGCGGTGACCGCGCGCCCGAGATCGGCTCCGAGGAGTGCAGG 10911  
Qy 3809 TTCGCGATCTCCGTGCGTACGCGCGGTGCGGTAGAGATGCTTCAATGAGTGCAGAAC 3868  
Db 10912 CGTTTCGCGCGGCGGAGTGAAGGCGCGGCGGCGGTCAATGAGAGGTCAATCTGTG 10971  
Qy 3869 CTTTCGCGCTGAGTGAAGGAGTCAAGAACGTTGCGTTTCCGCTGTCCACTCCGCG 3928  
Db 10972 GCGTGAATGTTCGAGGCGCGCGAGCGGAGGCGGAGGCGGTTCGCGGTGCGG-----GGGC 11026  
Qy 3929 GTTGAATCAACGACCGGCTGAGTGAAGTTCGAGTGAAGTTCGACACCGCGCGGATGTC 3988

Db 11027 GGTCAAGTTCGAGCGCGGATCGCGACAGACGAGGTGCGGACCGCGCGATGTC 11086  
 Qy 3989 CTCACCTGACACCGAGGAGCGGATCGGAGAGGAGGACCGCTTCGCGAGGACCTGCT 4048  
 Db 11087 GTCTCTCCCAATGAGCGGATCGGAGAGGAGGACCGCTTCGCGAGGACCTGCT 11146  
 Qy 4049 CTGTCTCAGCGCGAGCGGATCGGAGAGGAGGACCGCTTCGCGAGGACCTGCT 4108  
 Db 11147 GTCTCTCAGCGCGAGCGGATCGGAGAGGAGGACCGCTTCGCGAGGACCTGCT 11206  
 Qy 4109 GAAATGAGGCTGTGCGACGACCTTCCTCGCGAGGAGGACCGCTTCGCGAGGAC 4168  
 Db 11207 GAAATGAGGCTGTGCGACGACCTTCCTCGCGAGGAGGACCGCTTCGCGAGGAC 11266  
 Qy 4169 GATGCGGATGCGGATCGGAGAGGAGGACCGCTTCGCGAGGAGGACCGCTTCGCT 4228  
 Db 11267 GATGCGGATGCGGATCGGAGAGGAGGACCGCTTCGCGAGGAGGACCGCTTCGCT 11326  
 Qy 4229 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4288  
 Db 11327 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11386  
 Qy 4289 GTGTGATGACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4348  
 Db 11387 GTGTGATGACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11446  
 Qy 4349 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4408  
 Db 11447 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11506  
 Qy 4409 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4468  
 Db 11507 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11566  
 Qy 4469 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4528  
 Db 11567 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11626  
 Qy 4529 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4588  
 Db 11627 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11686  
 Qy 4589 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4648  
 Db 11687 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11746  
 Qy 4649 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4708  
 Db 11747 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11806  
 Qy 4709 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4768  
 Db 11807 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11866  
 Qy 4769 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4828  
 Db 11867 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11926  
 Qy 4829 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4888  
 Db 11927 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 11986  
 Qy 4889 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 4948  
 Db 11987 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 12046  
 Qy 4949 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 5008  
 Db 12047 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 12106  
 Qy 5009 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 5068  
 Db 12107 GGGCGGAGCGAGTGCAGCGGTGACCGCGGATCGCGAGGAGGACCGCTTCGCT 12166

Qy 5069 GATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 5099  
 Db 12167 GATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 12197  
 RESULT 3  
 US-09-861-289-3  
 ; Sequence 3, Application US/09861289  
 ; Patent No. US20020110897A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600.438US1  
 ; CURRENT APPLICATION NUMBER: US/09/861,289  
 ; CURRENT FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: 09/105,537  
 ; PRIOR FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 3  
 ; LENGTH: 13613  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-861-289-3  
 Query Match 14.6%; Score 1309.4; DB 3; Length 13613;  
 Best Local Similarity 67.7%; Pred. No. 8.4e-284;  
 Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;  
 Qy 2369 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2428  
 Db 10159 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10218  
 Qy 2429 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2488  
 Db 10219 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10278  
 Qy 2489 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2548  
 Db 10279 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10338  
 Qy 2549 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2608  
 Db 10339 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10398  
 Qy 2609 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2668  
 Db 10399 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10458  
 Qy 2669 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2728  
 Db 10459 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10518  
 Qy 2729 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2788  
 Db 10519 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10578  
 Qy 2789 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2848  
 Db 10579 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10638  
 Qy 2849 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2908  
 Db 10639 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10698  
 Qy 2909 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 2968  
 Db 10699 GGGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 10758



OY	2669	TTGGCCCCCCCCCGCAGAGATGATATGTGGTTAAGGCCCGAGCCGAGACCGCTGTCGCCCGCTTCG	3028
Db	10759	GAGGCACAGGCCCCCGGAGACAGATATGTAGCCGAACCCGAGGTTGATTCGCACTTCGCGCTCG	10818
OY	3029	GGCGGCAATCCGACAGAAAGTCCCGGAGGTTCTTCTCTGACCCGTTTGAAAGGCGCACGCTTG	3088
Db	10819	GGCGGCAAGCTGCTGTAAGCCGAGCCGACAGGTTCTTCGCGACGCGCGGAAGCGCGCTTCTTG	10878
OY	3089	CCGGTGGTGGCCTTGGTACTCGTCGTTGTTCAGCCCGGTACAGGAGGTTACGACCCGCGCG	3148
Db	10879	CCGGTGGTCTGTACTCTCGTCTGTGTAAGCCGTTAAGCGCGATAGCGAGGTGCGAATGGCGTCC	10938
OY	3149	AGGTTCACACAGGCCCGGCGCTGGCCGCGAGCGCTCTGTTCGTTGAGGGCGAAGCGTTGGTG	3208
Db	10939	AGGCCCCAGAGAGCCCGGCGCTGGCGCTTCACAGGATCGCTCGATTGACCGCAAGAGATTCTGG	10998
OY	3209	TAGACGGTCAAGGGCGAAGCCCCCGCGCGCGCGGTGCGACACCAACGCCCGAATACCCGGG	3268
Db	10999	TAGACGGTGGCCGCGAGCGCGTGGTCCGATTGCGCGCGCCAGGCTCCCGAGGCGGGG	11058
OY	3269	TTGGTCAACGAGTTCAGGCGCGCCGAGAGGTTACATCGCCGTCGGATTCTTCGACGGCACCC	3328
Db	11059	TTGGTGAAGCGCTCAAGCCCGCGCGAAGATACATCGCCGAGGGGTTGCCCGCGGTTATC	11118
OY	3329	TGCTGATATGACCGAGGTGAGGATCTCGTTGCGCGGACCAACCGACTCCGCTGTACTTG	3388
Db	11119	TGCTGATATGACCGGAAATGCGGATGCGCGCTTCGCGCTGAGGGCGGACCGGTTGTAGCCG	11178
OY	3389	GCACCTGTACCCGTTACCGAGAATGTACAGCGGAAATGTGACAGTCGCGGCGCGGCGTTAGAC	3448
Db	11179	GGCGCGGTCAACCGAGCGCAAGATGGCAGCGGAACATGTGAGTTGGGCGGGGTAGAGG	11238
OY	3449	CCGACGCTGTACCGGAAAGCGGAGCTCCCGCCGTACCGGACCTTCACAACGCGCCGCGCGCG	3508
Db	11239	CCGACGCTGTACCGGAAAGAGCGGCTTCGTGGCGAGCGCGCGTGCMAAGACGCGCGCTGT	11298
OY	3509	TTCAAGCGGCGTGAATCGATTGTGCTCCAGTACTTGCCCGCGCGGCTCTCTACCGCCGTG	3568
Db	11299	TGAGCGGGAACAGAGGTATGTTCTTCAGATACGCCCGCGCGGCGCGGCTTCGACCGCGGTG	11358
OY	3569	CCGAGCTGGGGAATGCAATGGAAGGTCGAGGAGCTGCGCAACCCGCGCGCGGTTCGACG	3628
Db	11359	CGAGACTCCGGAACCTGCCCCCAACAGGCGCGAGAGGCGCGCGAAGAGGTTCCCGATGACG	11418
OY	3629	TCGAACCGACGCGGCGATCTGTCTCCACAGCGGAGTGAACCGGCGCGCGCTGATGAGCGGTG	3688
Db	11419	CCGAGGTGTGGCGGCGCTCTCTTCAGCGGAGGTGAAGGGGCTGTTCCGTATGCGCACCGGG	11478
OY	3689	AGTTTCGCGAGCCCGGACCGGCTGTGCTTCGCGGAGTTCGATATGCGGCAACCGGTGAGCGG	3748
Db	11479	AGCCGGAAGAGGTGGCGGCGCGGTGTTCGCGCTCTGTTCGCGCGGCAAGAGCGCGCGGCG	11538
OY	3749	GTGACCTTCGGGCGCTGAGAGGCCCCGACCAACCGGCGCGCGCGGTGCGGAGTGCAGCGCGG	3808
Db	11539	GGAGAGGTTCTGGCCGACCGGCGGTGAACCGGCGCGCCCCCAATGCGCTTCGAGGATGGCGAG	11598
OY	3809	TCGCGGATCTCCGTCGCTGATACGAGCGGTGCGCCGTTAGAGTGTCTTCATGCACTGTGCAACC	3868
Db	11599	CGTTTCGGCGCGGGCGGTGGCGGAAAGCGCGGGGCGGTCACTGGAACGCTGCATGTGG	11658
OY	3869	CTTTCGCGGTCTGTGTGTGAGAGATCAAGAAACGTTGCGTTTCCGTTGTCCACTTCGCG	3928
Db	11659	GGGTGGAATGTCTGGGGGCGCGCGAGCGGGGCGGGGCGGTGTCCGCGGTGCG-----GCGC	11718
OY	3929	GTTTATCAACGCAACCGCTGTGTGTGTGGGAGATGCGATGACGTCGCAACCCGCGCGAATGTC	3988
Db	11714	GGTCAAGTTTCGCGGCGCGGATGCGCAAGACGACAGGTCTGGCGACCCGCGCGAATGTC	11778
OY	3989	CTCACTGAGCACCGAGAGGGCGGTTGCGAGAGGCGAGACCCGTTTCGAGAGCACTTCGCT	4048
Db	11774	GTCGTGCGCGAATGGCGGTGGCGGATTCGAGGGAACGACACGCGCGCGCGAGCGTTTCGCT	11838
OY	4049	CTGTGTCAAGCCGACCGGTGGCTTCGTCGCGGTGAGGGCGTCACTGTGTGGCAGCGCGGGA	4108

Db	11834	GTGGGGCAGCGGGGCGTGGCGGCTTGGCCCGCGGTACGGCTCCACTCTGTGGCAGCCCCGGGCA	11893
OY	4109	GAAGTAAAGGCTGTGGCAGCACTTCTCCGCGCGCAGAGATGCGCTGCAGCCGGTCAACGGTC	4168
Db	11894	GAAGTAAAGCGCGGGGTGTGCACGCGCTTCCGCGCTTCAAGAGACCTTCATAGACAGAGGTCCCGGTG	11953
OY	4169	GATGCCGCGTGTGGCGGTGCGCTTCCACAGATATATCATAGTCTGTAGTGTGCTCTCTCGTC	4228
Db	11954	GATGCCGCGTGTGGCGCTTCCGTTCGATCTTCAGAGATCAAGTACTGATGGTGTGTAAAGCCCGTG	12013
OY	4229	GGGCGGGAGGAGATGACCGGTGACGCGCGGATAGTCGCGCCAGCTCGCTGTGTAGAGCGC	4288
Db	12014	GCGGTCTGTGTCCGCGCAGACAGAGACGCGGGGAGGTCCGGAGGTGCTCGCGGTAAAGCCGCG	12073
OY	4289	GTGGTTGACCCCGGATGTGTGCTCCCTGTGTCTTCGCGGAAACGCGTCCAGGGAGTGAACCCCAT	4348
Db	12074	GTGGTTGCGCGCGGTCCGGTCCGGTCAATGACCTTCGGGAAACGCGTCCAGGGAGTGAAGCCCAT	12133
OY	4349	GGCGCGGGCGCACTCGCTCATCTTTCGCGGTGGTGGTGGCATCTGTGGTATCAACTTGTCCGG	4408
Db	12134	GGCGCGGGCGCGCTTCGCTCATCTTTCGCGGTGGTGGTGGCATCTGTGGTATCAACTTGTCCGG	12193
OY	4409	GCGGATGCGCGAGTGTGTGATGAGCGCGGATTCGGTTTGGGCGACAGCGCGCTGTCGGTGCAC	4468
Db	12194	GTCCAGACCCGAAATTTGTGAGAGGCGCGGATCCGGGCGCGAGAGTCCGACCTTCGTCCGTGAC	12253
OY	4469	CACCGCCCGCGCTTCGAGAGCGGTGACCGGCTTGTGTGGCGGTGAACTTGAACCTTCGCGC	4528
Db	12254	GACGGCGCGCGCTTCGAGAGCGGTGACCGGCTTGTGTGGCGGTGAACTTGAACCTTCGCGC	12313
OY	4529	GTCAACCGGATTCGCGCCACCGGAAGTCAACCCGTCGTGACAGCCAGGGCGGTGGCGCGTC	4588
Db	12314	GTCCCGCAGGCTGTGGCGGGCGGGCGCGCTGCACCGCGCAGGCCAGAGGCGGTGGCGCGTC	12373
OY	4589	GAAAGAAAGCTTGAACCTGTGTGTGGCGGGGAGATCTTTCGACAGCGCGCTTCACAGATGTCTGG	4648
Db	12374	GAATTTACAGCGCGCAGCGCGGTGTCTGTGTGGAGACCTTTCGCACTGTGTCCGCGCGCAGGG	12433
OY	4649	TCGCGCCCAACAGGTGTACGCGCGACGATTCGCGCGCGGTCTGTGGGGTGAACAGCGCGGACAC	4708
Db	12434	GCGGCGCCGAGAGGTGACGCGCGACGACGCGCGAGGTGGCGGGTGTACCGCGCGGCGCAC	12493
OY	4709	GTGGTCCGGGTTCGACCAAGACCGGTCCGCGGGTGTAGCTGCAGAAACAACCGGTGTAGTTC	4768
Db	12494	CTGGTCCGGGTTCGAGATTTCGCGGTGTCTCGGTGTGATGTGGCGAACAACCGGGTGAAGCC	12553
OY	4769	GAGCAACTGCGCGCGGTGGCGGGGTGGCGCGGAGAGTCAATTGAACGAGATATCATCTTCCGC	4828
Db	12554	GATCCAGCGCAGTGTGTGGCGGGGTGGCGGAGAGTCAATGACAGCATATCATCTTCCGC	12613
OY	4829	GGTGACGTTCACCGGCGCGCAGACCAAGTTCCAGGGGACGCGTGGCGTTTCAAGGTGGCAAT	4888
Db	12614	GGTAGAGCGCGCGCGCGTGGCGGAGAGAGTGAACCCGCGCGTGGCGTTGAGGTGCACAC	12673
OY	4889	GCAGTGCCCTTACCCCGACCAAGTGTGGGACCCCGGCGCTCGAATCTCCGCGACCAAGGGGCC	4948
Db	12674	GGCATATGCCCGGACCCCGCGAGACCCCGGCGAGCGCGCTCTTCACATCTTCGCGACAGAGCGGCG	12733
OY	4949	GCCGTTGTATGACCAAGTTGTGTTCACAGGGCCCATCTCAGGCGGGGCGCAGGAACCGCTGGCG	5008
Db	12734	GCCGTTGTATGACCAAGTTGTGTTCACAGGGCCCATCTCAGGCGGGGCGCAGGAACCGCTGGCG	12793
OY	5009	GTCCCGCATTCGTTCGAGCGCGCCCAAGTGCAGAGGGGTGCAGGAACGCTTCGCGACCGCGCA	5068
Db	12794	GTCCCATGCGGTGTGGGGCGCGCCCAAGAGAACCGCTGTCCAAAGCGCGGGCGCGCGAA	12853
OY	5069	GATCGCCAGATCGGTCCGATACGCGCTTCAGG	5099
Db	12854	GAATGCCAGGTCCGATTAAGGGCGCTTTTCACG	12884

```

US-09-860-846-3
; Sequence 3, Application US/09860846
; Patent No. US2002016472A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438051
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-3

```





Db 10939 AGACCCGAGAGCCGCGCTGCGCTCCAGGCTGCGCTCGGTGAGCCGGAAGGATTCTGTG 10998  
Qy 3209 TAGACGGTCAAGGCGAAGCCCGCGCGGCGCGGCGGCAACCAACCGCCCGATACCCGGG 3268  
Db 10999 TAGACGGTCAAGGCGAAGCCCGCGCGGCGCGGCGGCAACCAACCGCCCGATACCCGGG 11058  
Qy 3269 TTGGTCAACGGTTTCAAGCCCGCGAGAGGTACATCCGCTCGGTTTCTCCAGCGCAC 3328  
Db 11059 TTGGTGAAGCGGCTCAAGCCCGCGAGAGGTACATCCGCGAGGTTGCCCGCGGTATC 11118  
Qy 3329 TCGTCAATGACCGAAGTGAATCTGTGCGCGGCAACCACTCCGCTGTACTTG 3388  
Db 11119 TCGTCAATGACCGAAGTGAATCTGTGCGCGGCAACCACTCCGCTGTACTTG 11178  
Qy 3389 GCACTGTCAACCGGTACGAGAGTGAACAGCGGAACTGCACTGCGCGCGCGGTATAGC 3448  
Db 11179 GCGCCGGTCAACGAGAGCGAAGAGTGAACAGCGGAACTGCACTGCGCGCGGTATAGC 11238  
Qy 3449 CCGACGCTGTACCGGAGAGCGCGGCTCCCGCGGTACCGCAAGCTTCAACCGCGCGCGG 3508  
Db 11239 CCGACGCTGTACCGGAGAGCGCGGCTTCTGCGCGAGCGCGCGGTGAGAGCGCGCGCTGT 11298  
Qy 3509 TTCAAGCGCGCTGTACTGTGTTCTTCAAGTACTTCCCGCGCGCGGCTGTCTCACCGCGT 3568  
Db 11299 TTCAAGCGCGAGAGCGGCTGTCTTCAAGTACGCGCGCGCGCGGCTGTCTCACCGCGT 11358  
Qy 3569 CCGACGCTGTACCGGAGTGAATGAGAACTGAGCAAGCTGCGCGGAAACCGCGCGCGGTCAAG 3628  
Db 11359 CCGAGCTTCCGAGACTGCGCGGAACTGAGCAAGCTGCGCGGAAACCGCGCGGTCAAG 11418  
Qy 3629 TCGAACCGAGCGGCGCACTCTGCTCAACCGGCGTGAACCGCGAGGCGCGGTAGTGGAGT 3688  
Db 11419 CCGAGGCTGCGGCGCGGCTCTCTCAAGCGGCGTGAAGGCGCTGTGCGGTAGCGACG 11478  
Qy 3689 AGTTCCGCGAGCGGAGCCGCTGTCTTCCGCGGCGTGTGAGTCCGCGACCGGTGAAGCG 3748  
Db 11479 AGCGCGAGCGAGGCGGCGCGGCTGTCTTCCGCGGCTGTGAGTCCGCGACCGCGCGG 11538  
Qy 3749 GTGACCTCGGCGGCTGAGCGCGCGCGCAACCGCGCGCGCGCGCGGTGCGTCCGCG 3808  
Db 11539 GCGAGGCGTGTGCGCGAGCGGCGTGAACCGCGCGCGCGCGCGGTGCGTCCGCGCG 11598  
Qy 3809 TCCGCGATCTCGCTCGGTAAGCGCGGCTGCGGTAGAGTGTCTTCAATCGAGCGCAAC 3868  
Db 11599 CGTTCCGCGCGGCGCGGCGGTGCGGAGAGCGGCGGCGGTCAATCGGAGCGTCAATCGT 11658  
Qy 3869 CTTTCCGCGCTGTGTGTGCGAGATCAAGAACGTTGCGCTTTCGCTGTGCCACTCCG 3928  
Db 11659 GCGTGTGATGTCTGCGGCGCGCGAGCGCGGCGCGGCGGCGGTGCGCGGCTGCG 11713  
Qy 3929 GTTGAATCAACGACCGCTGTGTGTGCGAGTCCGATGACGTCGCAACCGCGCGATGTC 3988  
Db 11714 GGTCAATTCGCGCGCGCGGTGCGCGAGACCGAGAGGTCCGCGACCGCGCGAGTGTG 11773  
Qy 3989 CTCACTGGAACCGAGGCGCGGCTGCGGAGAGGCGGAGACCGCGTTCGCGCGAGCGACTCG 4048  
Db 11774 GTCTGTGCGCGATGTGCGGCTGCGGTGCGCGAGAGACCGCGCGCGCGCGGCTTTCG 11833  
Qy 4049 CTGTGTCAAGCGCGAGCGGTGCTTCCGTGCGGTAGGCGGTCAATCTGTGTGAGCGCGG 4108  
Db 11834 GTGCGCGAGCGGCGCGGTGCGGCTGCGCGGTGAGCGGTCAAGCTGCGGAGCGCGG 11893  
Qy 4109 GAAGTAGGCGGTGTGCGACGACTTCTTCCGCGCGAGAGTCCGTGACCGCGGTCAAG 4168  
Db 11894 GAAGTAGGCGCGGTGTGCGACGCTTCTTCCGCGCGAGAGTCCGTGACCGCGGTG 11953  
Qy 4169 GATGCGGCGGTGCGGTGCGGTGCGAGAGTCAAGTCACTGTGTAGTGTCTTCCGTG 4228  
Db 11954 GATGCGGCGGTGCGGTGCGGTGCGAGAGTCAAGTCACTGTGTAGTGTCTTCCGTG 12013  
Qy 4229 GCGCGGAGCGAGTGAACGCGTGAACCGCGGTGAGTGTGCGCGAGTCTTCCGTGTAGAG 4288  
Db 12014 GCGGTGCGGTGCGGAGCGCGGCGAGAGTGTGCGGAGTGTGCGGTGCGGTAGAGCG 12073

Qy 4289 GTGTTGACCGCGGTGTGCTCCCTGTGCTGCGGAGACGCGTGAAGGAGGTAGCCCAT 4348  
Db 12074 GTGTTGCGCGGTGCTCCGTGCTGAGTCACTTGGAGAACGCGTCAAGGAGGTGAGCCAT 12133  
Qy 4349 GCGCGCGCGCACTCTGCTCATCTTGTGCGGTGTGCGGAGTCCGTGACCACTTGTTCG 4408  
Db 12134 GCGCGCGCGCGCTGCTCATCTTGTGCGGTGTGCGGAGTCCGTGAGCGCGCGGTGAG 12193  
Qy 4409 GCGGATGCGAGTGTGCAATGCGCGGAGTCCGTGCGCGAGCAAGCGCGTGTGAG 4468  
Db 12194 GTGGAAGCGAGTGTGAGAGGCGCGGAGTCCGCGCGCGAGAGTGTGCGCGGTGAG 12253  
Qy 4469 CACCGCGCGCGCTGAGAGCGGTGACCGCGTGTGCGGTGAGTGTGAGTGTGAG 4528  
Db 12254 GACGCGCGCGCTTCAAGAGCGGTGAGCGCGTGTGCGGTGAGTGTGAGTGTGAG 12313  
Qy 4529 GTCAAGGATTCGCGCGAGCGGAGTCAACCGGTGCGGAGCGCGAGCGCGGTGAG 4588  
Db 12314 GTGCGGAGGCGTGTGCGCGCGCGCGCGCGGTGAGCGCGCGAGCGCGCGGTGAG 12373  
Qy 4589 GAAAGAGCTTGAAGTGTGCGGTGCGCGCGGATCTTTCGCGAGCGCTTCAAGTGTG 4648  
Db 12374 GAAGTACAGCGCGCGCGGTGCTGTGCGGAGCTTTCGCGAGCGGTGAGCGCGGTGAG 12433  
Qy 4649 TCGCGCGCGCAAGTGTGAGCGCGGAGTCCGCGCGGTGCGCGGTGAGCGCGGTGAG 4708  
Db 12434 GCGCGCGCGAGGTGAGCGCGGAGTCCGCGCGGTGAGCGCGGTGAGCGCGGTGAG 12493  
Qy 4709 GTGCTCGGCGTGAACAGCGCGGTGCGCGGAGTCAAGTGTGAGCGCGGTGAGTGTG 4768  
Db 12494 GTGCTCGGCGTGAAGTGTGCGCGGTGCGCGGAGTCAAGTGTGAGCGCGGTGAG 12553  
Qy 4769 GAGCGAGTTCGCGCGGTGCGCGGAGTCCGCGGAGTCAAGTGTGAGCGCGGTGAG 4828  
Db 12554 GATTCAGCGAGTGTGCGCGGAGTCCGCGGAGTCAAGTGTGAGCGCGGTGAG 12613  
Qy 4829 GTGAGTGAACCGCGCGCGGAGCAACGTTCAAGGCGAGCGGTGAGCGGTGAG 4888  
Db 12614 GTGAGGCGCGCGCGGTGCGCGGAGTCAAGTGTGAGCGCGGTGAGCGGTGAG 12673  
Qy 4889 GCGAGTTCGCGCGGAGTCCGCGGAGTCCGCGGAGTCAAGTGTGAGCGCGGTGAG 4948  
Db 12674 GCGAGTTCGCGCGGAGTCCGCGGAGTCCGCGGAGTCAAGTGTGAGCGCGGTGAG 12733  
Qy 4949 GCGGTGTGAGCGCGGAGTGTGTTCAAGGCGGAGTCCGCGGAGTCAAGTGTGAG 5008  
Db 12734 GCGGTGTGAGCGCGGAGTGTGTTCAAGGCGGAGTCCGCGGAGTCAAGTGTGAG 12793  
Qy 5009 GTGCGCGAGTGTGCGCGGAGTCCGCGGAGTCCGCGGAGTCAAGTGTGAG 5068  
Db 12794 GTGCGAGTGTGCGCGGAGTCCGCGGAGTCCGCGGAGTCAAGTGTGAG 12853  
Qy 5069 GATGCGCGAGTGTGCGGAGTCCGCGGAGTCCGCGGAGTCAAGTGTGAG 5099  
Db 12854 GATGCGCGAGTGTGCGGAGTCCGCGGAGTCCGCGGAGTCAAGTGTGAG 12884

RESULT 6  
US-10-271-889-46  
; Sequence 46, Application US/10271889  
; Publication No. US20030194784A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OR INVENTION: DNA Encoding Methymycin and Pikromycin  
; FILE REFERENCE: 600,582US1  
; CURRENT APPLICATION NUMBER: US/10/271,889  
; PRIOR APPLICATION NUMBER: 2002-10-15  
; PRIOR FILING DATE: 2001-05-18

```

; PRIOR APPLICATION NUMBER: US 09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-10-271-889-46

Query Match      14.6%  Score 1309.4;  DB 6;  Length 13613;
Best Local Similarity 67.7%;  Pred. No. 8.4e-284;
Matches 1850;  Conservative 0;  Mismatches 876;  Indels 5;  Gaps 1;

QY      2369  GGGCGGCTGACGCGCTTTCAGACGCGCGGATCAGGCCCCGATGCTGATCCACCCGCTCG
DB      10159  GGGGGGATCGGGGGGGGGTGGGGTCAAGCGAGAAAGCGGGGCTCTCCACAGCGCTCC
QY      2429  GCCAGCTCCGCTTTCAGATTGTTGAGCCGGGCGGCTCACCGACTGCTGAAACGCTGAGG
DB      10219  GCGCGCTCGCGCTTCAGCTGGTTCAGGCGGGCGGTGACGACTGATGAAAGCCGTCATG
QY      2489  AAGAAGCTGCTCCCGCGCTCGATGATGCTGCGGCGCGACGTCGACGAAAGTCGTGAGC
DB      10279  AAGAGTCTGCTCGCGCTGACGCGCGACCTCGCGCGCGCTGACGAAAGTCCCTGAC
QY      2549  ACCGAGTCAGAGCTCCGCTCGGGGTGACCGCGCGCGACGTCAGCGGGTGGCCCCGCG
DB      10339  ACCTCGGTGAGGAGGTGTGGGGGTCAAGCGCGCGCGGTGACGAGAGTCCCTGAC
QY      2609  AGCCCGGGAAACCGGCTCCCGGTACAGGTAGAGTCTCCGACAGGTGATCTGAC
DB      10399  AGCTCGGGAAAGCGGCTCCCGGTACAGGTACGTCGCGAGAGATCGACTTGAC
QY      2669  GCCACTGCGGATGGGGGTGGGGGATCGCTTCCGGCGGATCCGCAACGCTGGGCG
DB      10459  GCGACTGCGGGTGGCGGGTGGCGCGATGGTGGGGGCTTGATCCGAGCACTTCGCG
QY      2729  TCACACCCCGACGAGCTCTCCACGCGTACCCAGTCCGCTGTCATGCCGGGCTC
DB      10519  TCGGCCCCGATGCGCAGGCTCTTTCAGGCGGTAGTCGATGAGTGGAGTCCGGGGGTG
QY      2789  CGCTCGCGGCTAGTCAACGCGCGAGGCGCTCCCGCAACTCGCTCCGCTTCGCGC
DB      10579  CGCTCGGGAACCGCTCTCTCGAAGGCGTTGAGGGCTCTCGAGCTCGGCGCTCTCC
QY      2849  TCGGCAACCTGCGGTGTCGCGACCGGCTGATGCTCGGAGGAGGTGACGAGTCGAGC
DB      10639  TCGGCAAGCTGCGGTGTCGCGACCGGCTGATGCTCGGAGGAGGTGACGAGTCGAGC
QY      2909  GCGCGGTGCGGCTGACCTCGTTGAGTTCGCGGATGAGTCGACGAGTCGATGAGCGG
DB      10699  GTCTGCGCTTCGCGCGCTCGTTGAGTTCGCGGATGAGTCGACGAGTCGAGGAGCGG
QY      2969  TTGGCGCGCGCGCGGAGATGATGTTGAGGCGGAGCGGACCGAGCTGGCGCGCTTCG
DB      10759  GAGGACGCGCGCGGAGACGATGATGAGGAGCGGAGCTGATGAGCTCGGCTCG
QY      3029  GCGCGCATCCGCAAGATCCCGAGGTTCTTCGACCGGTCGAAAGGCGCAAGCTTCG
DB      10819  GCGCGCATCTGTCGAGAGCGGAGGTTCTTCGAGAGCGGCGGAGAGCGGCTTCCTG
QY      3089  CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
DB      10879  CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
QY      3149  AGGTCCCAAGGCGCGGCTGCGCGGCGAGCGTCTGTTGCTGAGGCGAAAGCGTTCG
DB      10939  AGGCGCGAGAGCGCGGCTGCGCGGCGAGCGTCTGTTGCTGAGGCGAAAGCGTTCG

```

```

QY      3209  TAGACGTCAGAGCGAGAGCCCGCGCGCGCTGCGACCAAGCCCCGATACCCGCG
DB      10999  TAGACGTCAGAGCGAGAGCCCGCGCGCGCTGCGAGCGCGCGAGCTCCGAGCGCGG
QY      3269  TTGGTCAAGCTTTCAGAGCGCGCGGAGAGGTACATGCGCTCGGGTTCGACCGGAC
DB      11059  TTGGTCAAGCTTTCAGAGCGCGCGGAGAGGTACATGCGCTCGGGTTCGACCGGATC
QY      3329  TCGTCAATGACCGAGGTGAGATCTGCTTTCGCGGAGCACCGACTCGGCTGATCTG
DB      11119  TCGTCAATGACCGAGGTGAGATCTGCTTTCGCGGAGCACCGACTCGGCTGATCTG
QY      3389  GCACCTGTCACCGCTACCGAGAAAGTACAGCGGAGATGACGTCGAGCGCGGTAGAGC
DB      11179  GCGCGCGTCAACGAGACGAGAGTGGCAGCGGAGATGACGTCGAGCGCGGTAGAGG
QY      3449  CCGACGCTGTCAGGAGAGCGGCGCTCCCGCTTACGCGAGCTTCACCGCGCGCGCG
DB      11239  CCGACGCTGTCAGGAGAGCGGCGCTCCGCTTACGCGAGCGCGCTTACGCGCGCGCT
QY      3509  TTGAGCGGCGCTGATCGGCTGCTCAGTACTTTCGCGCGCGCGCTTTCACCGCGCTG
DB      11299  TTGAGCGGAGAGAGGCTTCTTCAGTACGCGCGCGCGCGCTTTCACCGCGCTG
QY      3569  CCGAGCTGCGGATGCGATGAAACAGGTTCAGACGCTGCGGACCCGCGCGCTCAGG
DB      11359  CCGAGCTGCGGATGCGATGAAACAGGTTCAGACGCTGCGGACCCGCGCGCTCAGG
QY      3629  TCGAAGCGAGCGGATCTGCTCCACGCGGTGACGCGGAGCGCGGTGAGGCGGCTG
DB      11419  CCGAGCTGCGGATGCGATGAAACAGGTTCAGACGCTGCGGACCCGCGCGCTCAGG
QY      3689  AGTTGCGGAGCGGACCGCGCTGCTTCCGCGGTGCGATGCGGACCGGTCAGGCGG
DB      11479  AGCGGAGCGGAGCGGAGCGGCTGCTTCCGCGGTGCGGAGCGGAGCGGCGGCGG
QY      3749  GTGACTCGGCGTGAAGCGCGCGACCGAGCGCGCGGTGCGGCTGCGGCGG
DB      11539  GCGAGGCTGCGGCGGAGCGGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGGAG
QY      3809  TCGGCGATCTCGGTGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGC
DB      11599  CGTTGCGGCGGCGGCGGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      3869  CTTTGGCGCTGCTGCGGAGATGACGAGACCGTTCGCTTTCGCTTCCACTCGCGC
DB      11659  GCGTGAATGCTTGGGAGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      3929  GTTGAATGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB      11714  GTTGAATGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      3989  CTCACTGACACCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB      11774  GTTGAATGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      4049  CTGATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB      11834  GTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      4109  GAAATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB      11894  GAAATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      4169  GATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB      11954  GATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
QY      4229  GGGCGGAGCGAGTGCACGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
DB      12014  GGGCGGAGCGAGTGCACGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG

```

OY	4289	GTGGTGAACCGGGTTGGTCCTCCAGTCTGGGGGAAACGGCTCGAGGGAGGTGAAGCCCAT	4348
Db	12074	GTGGTTGGCGCGGGTTCCGGTCTGATGACCTCGGGAAAGCGGTGAGGGAGGTGAGCCCAT	12133
OY	4349	GGCGCGGGCGGCACTCGCTCATCTTGGCCGTTGGTGGTGGCCGATCTGGTGAACAACCTTGTCCGG	4408
Db	12134	GGCGGGCGGGCGGCTCTCGCATCTTGGGTTTGTCTCCGCGGGGGGGTGGCCGCGGGCGAG	12199
OY	4409	GCCGATGCGGAAGTTTGTGTCATGGCGCCGAACTCCGTTCCGCGACAGCGCCGTCTGGGTGAC	4468
Db	12194	GTCGAAGCGGAAGTTGTGTGAAGGGGCGGAATCCGGGGCGGAGAGTCCGCTGTCTGGTGAAC	12253
OY	4469	CACGGCCCGGCGCTCGAAGGGGGTGAACCGGCTTGTGTGGGTGGTGAAGCTGAACAACCTCGGC	4528
Db	12254	GACGGCGCGGCTCTCGAAGGGGTTGAAGCGCTTGTGTGGGTGGTGAAGCTGAAGACTTCGCGC	12313
OY	4529	GTCAACCGGATCCGCGCCACCGGACGTCACACCGCTGTGTGCAAGCCAGGGCGTGGGCGCGCTC	4588
Db	12314	GTCCCGGAGGGCTGGCGGGCGGGCGCGGCTGTGACCTGGCGCAACCGAGGGCGTGGCGGCGCTC	12373
OY	4589	GAGAAGAGCTTGAACCTGTGTGTGGGGGAGATCTTCCGAGGCGCTTCACAGTGTGG	4648
Db	12374	GAAATGACAGCGCGGCGGCGGCTGTCTGTGGGAGACTTTCGAGCTGTGGTCCGCGCGAGGG	12433
OY	4649	TCCGCCCCCAACAGTGTATCCCGCAAGATCCGGCGCGGCTGTGGGGGTGAACAAGCGCGGCAAC	4708
Db	12434	GCGGCCCCCAAGGGTGAAGCGCGCAAGAGAGGCGCAAGTGGGGGTGTGACCCGCGCGCGCAAC	12493
OY	4709	GTGGTCCGGGTGCAACGAACCGGTCGGCGGGGTGACGTCGCAGAAACACCGGTGTAGTGC	4768
Db	12494	CTGGTCCGGGTGCAAGTGTCCCGGTGTCTCCGGGTGATGTCCGCGAAGACCGGGTGAAGCC	12555
OY	4769	GAGCCAGCTCGCGCGGTCGGCGGTCGGCGCGGAAAGTCAATTGAACGACATGATCACTTCACC	4828
Db	12554	GATCCAGCGCATGTCGTGGGGGGTGGGGGAAAGTATGACGCGATGATCACTTCGCGC	12613
OY	4829	GGTACGTCACCGGCGCGGCAACAAGTTCAGAGGGAGACGTTGGCGTTGACGTTGGCGCAT	4888
Db	12614	GGTAGAGCGCGGCGGCGGTGGCGCAAGAGACTGAAGCCCGGCGTGTGGCTGTGACGTTGGCCAC	12673
OY	4889	GCAGTGCGGTAACCCCGCAACAGTGTGGGGGAGACCGCGGGCTCGAATCTCCGCAACGAGGGGCC	4948
Db	12674	GGCATGCGGAGACCCCGGCAAGCCCGGAGACGCGCTCTTCGAATCTCGGAGACGAGCGGGC	12733
OY	4949	GCCGTTGGTGAAGCCAGTGTGTGTTCAAGGGCCCACTCCAGCGGGGCAAGAACCGCTGGCGG	5008
Db	12734	GCGGTTGGAACAGCCACTGGCGTGTGAAGGGCCCGGTGAAGCGCTCTGAACGCTTGGCGGG	12793
OY	5009	GTCCCGCATGTCGCGCGCGGCCCAACGTGTCAAGGGGTGCAAGAACGCTTCGAGACCGCGAA	5068
Db	12794	GTCGATCGGGTGTGGGCGGCCCAACGAGAGACGGCTGATCGAAGCGGCGGGGCGCGGAA	12853
OY	5069	GATGCGCAAGTCCGTCGGTACGGGCTTCAGG	5099
Db	12854	GAATGCAAGGTCCGATGAAGGCGCTTTCACG	12884

[illegible]





Best Local Similarity 63.2%; Pred. No. 1.5e-268;  
Matches 1979; Conservative 0; Mismatches 1138; Indels 15; Gaps 4;

QY 2395 GGGGATCAGGCGCCGATGCTGATCCGACCGGTGGGCGACGTCCTTCGAGTGTGTAG 2454  
Db 55120 GCGAGAGAAACCCGCTGCTCCCGACGCAATCGGCGACGTCTCTCAATCTGTTCAG 59061  
QY 2455 CCGGCGGTGACAGCACTGTGCAAAACGTGAGGAAAGTCTGTCCTCCGCGTCCGATC 2514  
Db 59060 CCGCGGGGTACACCGCTGTGTGAAATCCGTCAAGAAATATTCTGTGCGCAGGCTTCGAG 59001  
QY 2515 GATGCTGCGGCGGACGTGACGAAGTCGTGACGACCGAGTGCAGGCTTCGCGGT 2574  
Db 59000 CACCGATCCGCGCTGTGTGACGAAGCTCCACCACTGTGCGCAAGCTGTGCGGAG 58941  
QY 2575 GACCGGCGCGGACGTAAGGCGGTGCGCCCGGACGCGCGGAAACCGGCTTCGCGTA 2634  
Db 58940 TACCGGTCCATCGAATGCGGTGCGCAACGCGCGGAAACCGGCTTCGCGTA 58881  
QY 2635 CAGGTAGACGTCTCCGACAGGTGATCTGACCGGCACTGTGCGATGCGGCGGCG 2694  
Db 58880 GAGGTAGACATCAACGAGGATATCCACTGTGACCGACACTGTGCGGCGGCGGCGG 58821  
QY 2695 CATGCTTTCGCGCGGATCCGCAACAGCTGCGGTGCAACCCGACGAGCTCTCA 2754  
Db 58820 CATGCTCTCGGCGGATACGCAAGAGCTCAAGTCTCCGCTTCGCGTCAACGAGCTG 58761  
QY 2755 CCGGTAAACCAAGTCCGTCTGCAATGCGCGGCGTCCGTGCGGCGGATGTCAGAAC 2814  
Db 58760 GGGGTGCGGTATGCAATGTCAGGCGGTGCGGTGCGGTGCGGTCTCTCCGCGAAC 58701  
QY 2815 GCGGAGGCGCTCTCCGCAACTCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 2874  
Db 58700 GTGCACTCTCGGCGGTGAGGCTCGGCGGCTGTCTCAAGGAGGCTTCGCTCCGCGG 58641  
QY 2875 GCTGTAATCTGCGGACGCTGCAAGTTCGAGCGGCGGTGCGGTGCGGTGCGGTG 2934  
Db 58640 GCTGTAATCTCTCCGCGGATGTAATGTCAGCGGCGGTGCGGTGCGGTGCGGTG 58581  
QY 2935 TTGCGGTGATGATTCGACAGGTCGATGAGGCGGTGCGGTGCGGTGCGGTGCG 2994  
Db 58580 CTGCGGTGATGATTCGACAGGTCGATGAGGCGGTGCGGTGCGGTGCGGTGCG 58521  
QY 2995 GTTGAAGCGGACGCGACGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 3054  
Db 58520 GCTGAAGCGGACGCGACGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 58461  
QY 3055 GTTCTTCTGACCGGTTCGAAAGCGCACGCTTTCGCGGTGCGGTGCGGTGCGGT 3114  
Db 58460 GTTCCGCGGTGACCGGTTCGAAAGCGCACGCTTTCGCGGTGCGGTGCGGTGCG 58401  
QY 3115 GTTCAAGCGGTGACGAGGTCGACGCGCGGAGGTCCACAGCGCGGCGGTGCG 3174  
Db 58400 GTTCAAGCGGTGACGAGGTCGACGCGCGGAGGTCCACAGCGCGGCGGTGCG 58341  
QY 3175 CAGCGTCTGTGCGGTGAGGCGGAGGCGGTGCGGTGAGGCGGTGAGGCGGCGG 3234  
Db 58340 GAGCTTCTGTCTCGTGAAGGCGGAGGCGGTGCGGTGAGGCGGTGAGGCGGCGG 58281  
QY 3235 GCGCGGTGCGGACCAAGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 3294  
Db 58280 GCGCGGTGCGGACCAAGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 58221  
QY 3295 GAGGTACATGCGGTGCGGTTCGCAAGCGCACCTGTGTGATGACGAGGTGAGATTC 3354  
Db 58220 GAGGTACATGCGGTGCGGTTCGCAAGCGCACCTGTGTGATGACGAGGTGAGATTC 58161  
QY 3355 GTTGAAGGAGACCAAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3414  
Db 58160 GTTGAAGGAGACCAAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 58101  
QY 3415 ACAGCGGAAATGATGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3474

Db 58100 ACAGCGGAAATGATGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 58041  
QY 3475 CCGCGGTACCGGACCTTCACAGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3534  
Db 58040 CCGGTGAGACCGGCGGTGAGAGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57981  
QY 3535 GTACTTTCGCGCGGCGGCTGCTGCAACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3594  
Db 57980 GTAGCGACCGGACGCTGCTGCTGCAACCGCGGTGCGGTGCGGTGCGGTGCGGT 57921  
QY 3595 GTGAGCAGCTGCGCGGACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3654  
Db 57920 GTGAGCAGCTGCGCGGACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57861  
QY 3655 CCGGAGTGAACCGGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3714  
Db 57860 CCGGAGTGAACCGGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57801  
QY 3715 TTGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3774  
Db 57800 CCGGATCTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57741  
QY 3775 CACGCGCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3834  
Db 57740 CCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57681  
QY 3835 CCGCGGTGAGGTGCTTCAATGAGTTCGCAACCTTTCGCGGTGCGGTGCGGTGCGGT 3893  
Db 57680 CTTCATGATGACGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57621  
QY 3894 CACGAAACGTTGGGT-----TTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 3945  
Db 57620 CCAAGCTTGTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57561  
QY 3946 TGGGTGTCGAGTTCGATGAGTTCGACACCGCGGTGCGGTGCGGTGCGGTGCGGT 4005  
Db 57560 GGGAGACCGCGCAACCGGATGAGTTCGACACCGGTGCGGTGCGGTGCGGTGCGGT 57501  
QY 4006 GCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4065  
Db 57500 AGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57441  
QY 4066 GTGCGTCCGTTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4125  
Db 57440 GCTGCGCGGACCGGTGAGGCTTCGATGAGGCGGTGCGGTGCGGTGCGGTGCGGT 57381  
QY 4126 CGACCTTCTCCGCGGACGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4185  
Db 57380 CGACCTTCTCCGCGGACGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57321  
QY 4186 CGTCAACAGATGATTCAGTACTGATGTTGCTCTCTGTCGCGGCGGAGCGATGCA 4245  
Db 57320 CGTCAACAGATGATTCAGTACTGATGTTGCTCTCTGTCGCGGCGGAGCGATGCA 57261  
QY 4246 CCGTGAACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4305  
Db 57260 CCGTGAACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57201  
QY 4306 GCTCCCTGCTCTCGCGGACGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4365  
Db 57200 GCGCGGCGGACCTTCGCGGAAACGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57141  
QY 4366 TCATCTTTCGCTGTTGTCGCGGCGGCTGCGGTGCGGTGCGGTGCGGTGCGGT 57081  
Db 57140 TCATCTTTCGCTGTTGTCGCGGCGGCTGCGGTGCGGTGCGGTGCGGTGCGGT 57021  
QY 4426 GGAAGGCGGATTCGCTTTCGCGGACGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4485  
Db 57080 CCAAGGAGGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 57021  
QY 4486 AGCGGTGACCGCTTTCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4545  
Db 57020 AGCGGTGACCGCTTTCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 56961





LOCATION: (15643) .. (17466)	LOCATION: (15643) .. (17466)
OTHER INFORMATION: ORF28	OTHER INFORMATION: ORF28
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((17522) .. (18895))	LOCATION: Complement ((17522) .. (18895))
OTHER INFORMATION: ORF27	OTHER INFORMATION: ORF27
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (19063) .. (20229)	LOCATION: (19063) .. (20229)
OTHER INFORMATION: ORF26	OTHER INFORMATION: ORF26
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((20307) .. (21743))	LOCATION: Complement ((20307) .. (21743))
OTHER INFORMATION: ORF25	OTHER INFORMATION: ORF25
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((21733) .. (22527))	LOCATION: Complement ((21733) .. (22527))
OTHER INFORMATION: ORF24	OTHER INFORMATION: ORF24
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((22534) .. (23571))	LOCATION: Complement ((22534) .. (23571))
OTHER INFORMATION: ORF23	OTHER INFORMATION: ORF23
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((23555) .. (24463))	LOCATION: Complement ((23555) .. (24463))
OTHER INFORMATION: ORF22	OTHER INFORMATION: ORF22
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((24460) .. (25650))	LOCATION: Complement ((24460) .. (25650))
OTHER INFORMATION: ORF21	OTHER INFORMATION: ORF21
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((25647) .. (26105))	LOCATION: Complement ((25647) .. (26105))
OTHER INFORMATION: ORF20	OTHER INFORMATION: ORF20
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: Complement ((26180) .. (27391))	LOCATION: Complement ((26180) .. (27391))
OTHER INFORMATION: ORF19	OTHER INFORMATION: ORF19
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (27937) .. (28983)	LOCATION: (27937) .. (28983)
OTHER INFORMATION: ORF18	OTHER INFORMATION: ORF18
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (29244) .. (42779)	LOCATION: (29244) .. (42779)
OTHER INFORMATION: ORF1	OTHER INFORMATION: ORF1
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (42823) .. (48657)	LOCATION: (42823) .. (48657)
OTHER INFORMATION: ORF2	OTHER INFORMATION: ORF2
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (48712) .. (59802)	LOCATION: (48712) .. (59802)
OTHER INFORMATION: ORF3	OTHER INFORMATION: ORF3
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (59850) .. (64556)	LOCATION: (59850) .. (64556)
OTHER INFORMATION: ORF4	OTHER INFORMATION: ORF4
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (64687) .. (70365)	LOCATION: (64687) .. (70365)
OTHER INFORMATION: ORF5	OTHER INFORMATION: ORF5
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (70365) .. (71078)	LOCATION: (70365) .. (71078)
OTHER INFORMATION: ORF6	OTHER INFORMATION: ORF6
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (71113) .. (72360)	LOCATION: (71113) .. (72360)
OTHER INFORMATION: ORF7	OTHER INFORMATION: ORF7
FEATURE:	FEATURE:
NAME/KEY: CDS	NAME/KEY: CDS
LOCATION: (72400) .. (73665)	LOCATION: (72400) .. (73665)

```

1  OTHER INFORMATION: ORP8
2  FEATURE:
3  NAME/KEY: CDS
4  LOCATION: (73694)..(75043)
5  OTHER INFORMATION: ORP9
6  FEATURE:
7  NAME/KEY: CDS
8  LOCATION: Complement((75899)..(76570))
9  OTHER INFORMATION: ORP10
10 FEATURE:
11 NAME/KEY: CDS
12 LOCATION: Complement((76602)..(77765))
13 OTHER INFORMATION: ORP11
14 FEATURE:
15 NAME/KEY: CDS
16 LOCATION: (78039)..(79313)
17 OTHER INFORMATION: ORP12
18 FEATURE:
19 NAME/KEY: CDS
20 LOCATION: Complement((79391)..(81052))
21 OTHER INFORMATION: ORP13
22 FEATURE:
23 NAME/KEY: CDS
24 LOCATION: (81541)..(82356)
25 OTHER INFORMATION: ORP14
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: (82760)..(83362)
29 OTHER INFORMATION: ORP15
30 FEATURE:
31 NAME/KEY: CDS
32 LOCATION: Complement((83495)..(84142))
33 OTHER INFORMATION: ORP16
34 FEATURE:
35 NAME/KEY: CDS
36 LOCATION: (84329)..(84428)
37 OTHER INFORMATION: ORP17 (Fragment)
38 JS10-229-148B-1

```

Query Match	13.5%	Score 1203.2;	DB 7;	Length 84428;
Best Local Similarity	65.5%	Pred. No. 5e-260;		
Matches 1778; Conservative	0;	Mismatches 933;	Indels 5;	Gaps 1;

QY	2387	AAGACGGCGGGGAAATCAAGACCCCGAATGCTGAATCCCAACCCGTCGGGAGAGTCCCGTTCCAGT	244.6
Db	3332	CAGGCCCTAATCGAATAAAGCCGCGCTGGTCCGCCACCCGTCGGCGGTAATCACTTCAGAC	329.1
QY	2447	TGGTTGAGCCGGGCGGTGATCCGACTGGTTCGAAACCGTCGAGGAAGAACTGCTCCCGCGC	250.6
Db	3392	TGGTTCAAGCCGTGCGGTGAAGACCTGGATCGAAACCGTCGATGAAGTACTGCTCCCGCTCG	335.1
QY	2507	TGCGGATCGATGCTGCGCGGCCGACGTTACGAAAGTCGTGAGCGAACCGAGTGCAGGCTCCG	256.6
Db	3352	GCGGAGGAATCTGCCACCGCTGGTCAAAAGCGCTCAACACTGGGTCAAGGTTGTT	341.1
QY	2567	TCCGGGGGTGAACCCGGCCGCGAGAGTAGCGGGTGGCCCCCGCAGCCCGGGGAAACCGGCG	262.6
Db	3412	CCGGGGGAGACTTGTGCGGATGCTGTAGCGGTCCGCTCCGGGCAAGTCCCGGAAACCGGCG	347.1
QY	2627	TCCCGGTACAGGTAGAGAGTCTCCGAGGAGAGGTCCGATCTGCACCGCACTCTGCGATYGGGCG	266.6
Db	3472	TCCGGGTAGAGGTAGAGTGTGCGCCAGCAGATTCGACTGCGACTTGGCGGATYGGGCG	353.1
QY	2687	GTCGGGCGCAGTCGTTTCGGGCGCGGATCCGGAACAGCTGGGCGGTGACCAACCCGACGAGG	274.6
Db	3532	GTCGGGCGCAGTCGTTTCGGGCGCAATTCGGAACGAGTTCCGCGTGGGCTCCGGCCGAGAGG	359.1
QY	2747	CTCTCCAAACCGCTAACCCAGAGTGGTCTGCATGCCCGGGTCCGCTTCGCGCGCGCTAGTCG	280.6
Db	3592	CTGTTCAAGCGCGTACCCGTAGTGAACGTGAGGGTGGGTGTTCTTGCGCGGCACTTCTCC	365.1
QY	2807	ACGAACCGGGCGAGGCCCTTCCGCAACTCGATCCGTTGCGCTTCGAGCAACTTCGCGGTG	286.6

Db 3652 CGAAGCCGGTCAGCGCTTCTCGAGGTCGCGCGCTCGGCACCGGACAGAGAGCCGTCG 3711  
Qy 2867 TCCGCAACGGTGAATGCTCTGCGCGGACGGTGAACGAAGTCAGCGCGCGGGTGGGCTGAC 2926  
Db 3712 GCGCGCGCGCTGAATGCTCTCAACGAGGATGACGAATGTCAGTGGTCCGTCGCGCGCGCT 3771  
Qy 2927 TCGTTGAGTTCCGCGATGAAGTCGACCAAGTCAGTGAAGCGGTTGAGCCCGCCCGGACG 2986  
Db 3772 GCGTTCAACTCGCGGATGAAGTCAGCAAGTCAGAGCGCGCGCGCCCGACCGCGACG 3831  
Qy 2987 ATGATGATGTTGAAGCCGAGCCGAGCCGCGTCCCGCTTGGCGCGCATCCGACAGAG 3046  
Db 3832 ACATGTAATGTTCAAGCCGAGTTTCAACCGTTTCGCGCGTGGGAAACCAAGCCGCTGAG 3891  
Qy 3047 TCCCGAGATTTCTTCTGACCCGTTCCAGGCGCCACGCTTGGCGGTTGATCCGTTGAC 3106  
Db 3892 CGCTCCAGGTTGGCTTTCACCCGTCGAAAGCGCTTCTTCTGATGTTGCGCGAT 3951  
Qy 3107 TCGTGTGTTGACCGCGTACAGAGAGTTACGACCGCGCCGAGGTCCACAGCGCCGAGC 3166  
Db 3952 TCCGCAATGTTCAACCGGTACAGGAGGTCCGCAACCGGTGACAGCGCGCACAGCGCGCC 4011  
Qy 3167 TGGCGCGCGACGCTCTGTTGCTGAAGGCGAAGCGCTTGTGTAGACGTCAGGCGAG 3226  
Db 4012 TGCTTTCCAGCGTCCGGTCGGTAAGCGCGAAGCGCTTGTGTACAAAGTTAGCGGAAAT 4071  
Qy 3227 CCGCGCGCGGCGGCGTGGACCAACAGCCCGCATCCCGGTTGTTCAAGTTCCAG 3286  
Db 4072 CCGCGCGCGGCGGCGCTCCGAAACAGAGGAGCCAGCCCGGATTTGTAGCGGCTCCAG 4131  
Qy 3287 CCGCGCGAGAGATACATCGCGCGTGGCTTCCGACGCGACCTGTCGATGACGAGGTG 3346  
Db 4132 CCGCGCGACAGTACATGCGCGTGGGTTGCTCGTTGACATGTCGTGATAGAGCGCGCG 4191  
Qy 3347 AGGATCTGTTGCGCGGACCAACGCACTCCGCTGTGATCTTGCACTGTCAACCGTACG 3406  
Db 4192 AACATCTGTTGCGCTCGTCCAGGCGGAGTGAATGTAGCGGCGCCCGGTACCCGTACG 4251  
Qy 3407 CAGAAGTGAACAGCGGAATGCAAGTCGCGCGCGGTTAGACCGGACGCTGATAGGAG 3466  
Db 4252 CAGAAGTGAACAGGAATGCAAGTCGCGCGCGGATGACCGCACGGAATGCGGAGC 4311  
Qy 3467 GCGGCGTCCCGCGTACCGGACGCTCCAAACGCGCGCGCGGCTTACGCGCGCTGATCGTG 3526  
Db 4312 ACCGCTTGTGTTGAAGCGCGCTGAAAGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 4371  
Qy 3527 TTGCTCAGTACTTGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3586  
Db 4372 TTGCTCAGTACTTGCGCGAGGCGGCTTCTCAACGCGCGCACTGCGGACCAACG 4431  
Qy 3587 TCGAACAGTGAAGCTGCGCGAACCCGCGCGCGGTCCAGGTGAAACCGACGCGCATC 3646  
Db 4432 CCGAAGAGTCAAGAGACGCGGAGGACAGATCGGTGACTCTTGAATGAGCGGCGC 4491  
Qy 3647 TGCTCAACCGGAGTGAAGCGGACGCGGCTGATGAGGCGGTTGCGGCGAGCGGAC 3706  
Db 4492 TTTTTCAGCGAGTGAAGGAGCGCGCGCGGTGATCCGCGCGAGCGCTTCAACCAATGCGCG 4551  
Qy 3707 GCGTGTGTTCCGCGCGGTGTCGAGTCGAGCCGAGTGAAGCGGTTGACCTCGCGCTGAGC 3766  
Db 4552 GCGGCGTGGCGCGCATCCGCGTCCGTCATATGCGCGGAGTGAACAAGTTGCGCGGAGA 4611  
Qy 3767 GCGCGCACACGCGCGCGCGGTGCGCGGTGCGGTGCGCGGTCCGCGCATCTCGCTCGGT 3826  
Db 4612 GCGTGAAGCGCGCGCGGATCACTGCGCGGAGGATGACACGCTAAGTGTTCGC 4671  
Qy 3827 ACAGCGGTGCGCGTGAAGTGTCTTCAATGACGTGCAACCTTCTGCGCGCTGATGAG 3886  
Db 4672 ACCGCTTCTTCTCAACATGAGTATCCCAACCTTCTTCTCAATGCGAATGCTTGCATG 4731  
Qy 3887 CGAGATCA-----CGAACGTTGCGTTTCGCTTGTCCCACTCGCGCTGATCAACGA 3941  
Db 4732 AGAAGGCGGTGCGGTGATCACTGAGTGCCTGACGCGAGCGGCGGTCAAGCTCGGCG 4791

Qy 3942 CCGCTGATGATGCGAGTCCGATGACGTCCACACCCCGCGATGTTCTTCACTGACACC 4001  
Db 4792 CCGCGGTTGCGCGCGCACCGGACGATGTCATACCCGCGGATATCTTCACTGAGAGACC 4851  
Qy 4002 GAGGCGCGGTCCGAGAGGTCGAGACCTCTTCCGCGGACCACTTCGCTCTGTTCACGCCG 4061  
Db 4852 GTGACCCCGATCCGACAGGCGGATGACCCGCGCGAGAGGCGCTCGGTGTCGAGAGGTG 4911  
Qy 4062 AGCGGTGCTCCGTCGCGTGAAGGCGTCACTGATGACAGCGGCGGAGAGATGAGGCTGT 4121  
Db 4912 GCGTTCCGACGGAACCGGTACGCTTCAGTTGGTTGCGAGGCGCGCGAGAGTGAAGGCTGG 4971  
Qy 4122 GCGACGACTTCTCCGCGCGCAGAGATCCCTGACCGCGTCACTGTCATGTCGCGTGG 4181  
Db 4972 GCGACGACTTCTCCGCGGAGAGAGCGAGCAGATCCGATGAGACCGCGTGAAC 5031  
Qy 4182 GTGCGCTCAACAGATGATCACTGATGTTGCTCTCTCTGTCGCGCGGAGAGAG 4241  
Db 5032 TCCCTGCTGATGTCACGACAGATCACTGATGTTGCTCTCTCTGTCGATGAAATGCG 5091  
Qy 4242 TGACCGGTGAACGCGGCTGATGTCGCGAGCTGCTGCTGATGAGCGGCTGTAACCGG 4301  
Db 5092 AAGACGCGCACACCGGTACTCCGAGAGTTCCGTGCGGTATGCTCATAGTTGCTTGG 5151  
Qy 4302 TTGCTCTCCCTGCTCTCGCGGAAACGCTGAGAGGATGAGCCCATGAGCCGCGCGAC 4361  
Db 5152 TTGTCGCTGACGACTTCTTCAACAGCTCGAGGAGCGTCAAGCCCATGCGCGAGGCT 5211  
Qy 4362 TCGCTCATCTTCCGCTTGTGTCGCTGATCCGTCACACTTGTCCCGGCGGATGCGAG 4421  
Db 5212 TCGCTCATCTTCCGATTTGTCCTCCCGCGGAACTGACTTCTTCCAGCGGAGTCCAG 5271  
Qy 4422 TTGTCATGTCGCGGATCCGTTGCGGACGAGCGCGCTGTCGAGTACACCGCGCGCGCC 4481  
Db 5272 TTGTCATGTCGCGGATCCGTTGCGGACGAGCTGTCGATGTCGATGTCGATGTCGATG 5331  
Qy 4482 TCGAAGCGGTGACCGGCTTGTGTCGCTGAGAGCTGAAACCTTGGGCTCAACCGATCCG 4541  
Db 5332 TCGAAGGAAATGACGACCTTCTGTCGAGTGAAGCTGAATCTCGGCGTCCGACCGG 5391  
Qy 4542 CCGACCGGACGTCACCGCTGTCGACCGCGCGGCTGAGCGGCTGAGAGAAAGCTTG 4601  
Db 5392 CCGACGAGCTTCTCTGCGAGTGTCTGCGATGCGTGCCTGCGGCGGAGAAACGCGCG 5451  
Qy 4602 ACCGTGATGTCGCGGCGGATCTTTCGACGCGCTCAACAGTGTGATCGGCGCCACAG 4661  
Db 5452 ATGCGGCTTCCGCGCGGACGCTCTTCAAGCGCTGACATCAACGCGCTTACCCACAG 5511  
Qy 4662 TGTACGCGAGATCGCGCGGCTGTCGCGGAGTGAACGAGCGCGGACGATGTCGCGGTG 4721  
Db 5512 TGTACCGGAAATGCGGAGTGTGCGGAGTGTGAGGCGCGGACCGCGCTCGGATCC 5571  
Qy 4722 ACCAAGCGGTGCGCGGCTGACCTTCCGACAAACCGGTGATGATCCGAGCGACTCGCC 4781  
Db 5572 ACAATCGGTGATGATGATCGATCGGAGAAACCGGCTTCCAGTCCGAGCACCGCACT 5631  
Qy 4782 GCGTGCCTGCGGCGCGGAGGATGATGAGCGGATGATGATGATGATGATGATGATGATGAT 4841  
Db 5632 GCGTGCCTGCGGCGCGGAGGATGATGAGCGGCGGAGGATGATGATGATGATGATGATGAT 5691  
Qy 4842 GCGCGACGACCAATTCAGGAGGACGATGCGTTGCAAGTGTGATGATGATGATGATGATGAT 4901  
Db 5692 GCGTGCCTGCGGCGCGGAGGATGATGAGCGGCGGAGGATGATGATGATGATGATGATGAT 5751  
Qy 4902 CCGACGAGTGGCGGACCGCGGCTTCCGATCTCCGACCAAGGCGCGCGCTTGTGAGC 4961  
Db 5752 CCGGCGAGATGGCGGACCGCGCTTCCGATCTTGGAGTCAAGGCGCGCGCTTGTGAGC 5811  
Qy 4962 CAGTTGTTTCAAGGCGGACCTTCCAGGCGGCGGAGGACCGCTGCGGCTGCGATGCTG 5021  
Db 5812 CACTGTGTCGAGGCGGAGGCTGAGCGGCTGAGCAACCGGAGGAGGATTC 5871



## RESULT 11

US-10-201-365-22/c

Publication No. US20030148469A1

Publication No. US20030148469A1

; GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary

APPLICANT: BETLACH, Melanie C.

APPLICANT: BETLACH, MARY

APPLICANT: BETLACH, Mary

APPLICANT: MCDANIEL

APPLICANT: TANG, IJ  
APPLICANT: MCDANIEL, ROBERT

APPLICANT: TANG, LI

; TITLE OF INVENTION: COMBINATORIAL POLYKET

TITLE OF INVENTION: COMBINATORIAL  
TITLE OF INVENTION: ;

1. TITLE OF INVENTION: PKS GENE CLUSTER A

FILE REFERENCE: 300622002103

FILE REFERENCE: 300622002103

! CURRENT APPLICATION NUMBER: U

CURRENT APPLICATION NUMBER: 3003-07

**CURRENT FILING DATE: 2002-**

PRIOR APPLICA

PRIOR APPLICATION

PRIOR FILING

PRIOR FILING DATE: 1998-08-28  
APPLICATION NUMBER: IIS 09/07

**PRIOR APPLICATION**

```

!
! PRIOR FILING DATE: 1998-05-06
!
! NUMBER OF SEQ ID NOS: 32
!
! SOFTWARE: PatentIn Ver. 2.0
!
! SEQ ID NO: 22
!
! LENGTH: 1565
!
! TYPE: DNA
!
! ORGANISM: Streptomyces venezuelae
!
US-10-201-365-22

```

Query Match	Similarity	8.2%	Score 729.4	DB 6	Length 1563
Beet Local	Similarity	68.1%	Pred. No. 1e-153	Matches 1015	Conservative
2369	GAGCGAGTTCGACGCGGTTCCAGACGCGCGGATCAGGCGCCGATGCTGATTCGCCACCGTGC	2428			
1530	GAGGGGAGATCGGGGCGGGGTGCGGGATCAGGCGCAGAAAGCGCGGGCGCTCTCCACGCGCTGC	1473			
2429	GCGAGTCTCCGTTCCAGATTGGTTTGAACCGGGCGGTCACCGAATTGGTCCAAACCTGTCAGG	2488			
1470	GCGGGGTCCCGCTCCACGACTGGTTCCAGCGGGGCGGTGACCACTGATTCGAAAGCGTCTCATG	1411			
2489	AAGAACTCGTCCCGCGGCTGCGGATCGAATGCTGCGGCGCCGACGTACAGAAATGCTGAGG	2544			
1410	AAGTACTCGTCCCGCTCCAGACGCGCGCACCTCGCGCGCGGTCTCAGCAGAAATCTCTGACG	1351			
2549	ACCGAGTGCAGGCTCCGATCCGAGGTGACCCCGCGCGGACGACGTAGCGGTTGCCCCCGCGC	2601			
1350	ACCTCGATGAGGAGAGGTGTCGGGGGATCAGCGGGCCGCGCATGTAGCGGGTCTCCGCGCTCC	1291			
2609	AAGCCCGGGGAAACCGGCGTCCCGGTACAGTGTAGAGTCTTCGAGCGAGGTGCAATCTGACGC	2661			
1290	AAGTCCGGGAGAGCGGCTCGCGGTGACAGTACAGTCCGCCGAGGAGATCGACTTGACCTGACC	1231			
2669	GCCACTTGCAGATGCGGCGGTGCGGCGCATCGTTTCGCGCGGATTCGCCAACAGCTGGGCG	2721			
1230	GCGACTGCGGGGTGCGCGGTGCGGCGCATGTGTGCGGGGCTTGATTCGCGACAGTTCGCGG	1171			
2729	TGCAACACCCCGAGACGAGGCTTCTCCACCGGTACACCAAGTCCGATCTGCAATGCCCCGGGTC	2781			
1170	TGCGACCCCGATGTGCGAGGCTGTTCACAGGCGGTGAGCCCTGATGTGATGTGAGGTCCGGGGGTG	1111			
2789	CGCTCGCGGAGCGTGAATCGACGAAACCGGGGAGAGGCGCTCCCGCAATCTCGATCTCGCTCC	2841			
1110	CGCTTCGCGGACCGGCTCTCTCGAAGGCGGTGAGGGCTCTCTGAGAGCTTCGCGCGGCTCTCC	1051			
2849	TCGCAACAATCTGCGCTGTTCCCGACGCTGTAGTCTCTCGCGACGCGTGAAGAAATCGACG	2901			
1050	TGCGGAGAGCTTGCCTGCTGTCACGCGCGCGTGTAGTCTCTCGCAATGTTTGACGAAGTGCATC	991			
2909	GCGCGGTGCGGCTGCACTCGTTGAATTCGGCGAATGAATGTCACAGAGTGCATGAGACGG	2961			
990	GTTCTGCGCTTCGCGCGCGCTGTGAGGTGCGGCGATGAATGTCACAGAGTGCACAGAGCG	931			

```

Oy      2665  TTGGCCCCCCCCCAGAGATGATGTGGTTTAAAGCCGAGCCGGAACCGGTGGCCGCGTTGG 3028
Db      930  GAGGCAAGGCCCCGGGAGACAGATGTAGCCGAAAGCCGAGGTTTATCGGCGACTGCGCTCG 871
Oy      3029  GCGGCGCATCCGAGAGAAAGTCCCGGAGGGTCTTTCCTCTGACCCGTTTCGAAAGCGCAACGCTTG 3088
Db      870  GCGGCGCATCTGTGTGAAGCCGCGCGAGGTTCTTCGCGAGACGGGCGGAAAGGCGGCGCTTCTTG 811
Oy      3089  CCGGTGGTGGCTGTGTACTGTCTGTGTTTCAAGCCCGTACAGAGGAGTTACGACCCGCGCG 3148
Db      810  CCGGTGGTGTCTGTGTACTCTCTGTCTGTGTGAGCGCTGTAGAGCGAGAGTGGCGGATGGCTGC 751
Oy      3149  AGGTCCCAACAGCCGGGCGTGGCGGCGAGGCTGTGTTCCGAGAGGCGAAAGGCGCTTGTGT 3208
Db      750  AGGCCCCAGAAAGCCGGGCGTGGCGCTTCCAGGGTGGCGCTCGGTGACCGGAAAGGATTTGTGT 691
Oy      3209  TAGACGGGTCAAGAGGCGAAAGCCCGCGCGGCGCGTGGCGACACCAACCGCCCGATACCCGGG 3268
Db      650  TAGACGGGTGGGCCCGCACGCGCGTGGTCCGTGTGGCGTGTCCGCGACAGGCTCCGAGGCGCGGGG 631
Oy      3269  TTGGTCAACGGTTTCAAGGCGCGCGCGGAGAGTTACATCGCCGTCGTGGGTTCTTCGACGGCAC 3328
Db      630  TTGGGTAGCGGCTCAAGGCGCGCGCGGAGAAATATCCCGAGAGGTTTGGCCCGCGAGGTATC 571
Oy      3329  TCGTGCATGACCGGAGGTGGAGGATCTCGTTGGCGGGGACACACGACTCCGCGCTGTACTTGG 3388
Db      570  TCGTGCATGACCGGACCGGACATGAGCGTTGCGCGGCTGTAGAGGCGGACGGGTCGTATAGCG 511
Oy      3389  GCACCTGTGACCCGTTACGCAAGAGTTGACAGCGGAGCAATGCAAGTCGGGCGCGGAGTGAAGC 3448
Db      510  GCGCGGTGTCAACCGGACCGCAAAATGGGAGCGGAAACATGCAAGTCCGGGCGGGGTATAGAG 451
Oy      3449  CCGACGCTGTACGGGAAAGCGGAGCTCCCGCGGTACCGAGCTTCCAAACGCGCCGCGCGG 3508
Db      450  CCGACGCTGTACGGGAAAGCGGAGCTTCTGTGGCGAGCGCGCGCTGAAAGACCGCGCGCTGT 391
Oy      3509  TTCAAGGCGGCTGATCTGTGTGTCTCAAGTACATTGGCCCGCGCGGCGCGTGTCCACGCGCGGT 3568
Db      390  TCGACCGGAGACAGGAGTGTCTTCCAGTACGCGCCCGGCGGAGCGGTTTCAACGCGCGGTGT 331
Oy      3569  CCGACGCTCGGAGATCGATTCGACAGAGGTGAGAGACTGCGCGAAACCCGCGCGGTTCAGG 3628
Db      330  CGGAGCTTCGGGACCTGTCCCGCAAGAGGCGCGAGAGGCGCGCGAAAGGCGTCCCGGTTCAGG 271
Oy      3629  TCGAACCGACGCGCGCATGTGCTCCACCGGAGTGAACGGCGAGGCGCGGTAGTGGCGGTG 3688
Db      270  CCGAGGTGTGGCGGGGCTCTCTCCAGCGGAGTGAAGGAGGAGCTGTGTTCCGTACGCAACGCG 211
Oy      3689  AGTTCCGCGGAGCGCGGACCGCGCTGTCTGTTCGCGCGGTGTGAGATGGCGGACCGGTGAGGGCG 3748
Db      210  AGCCCGGAGAGAGGTGGCGGCGGCGGTCTGTTCGCGCTCTGTGTGGCGCGGACAGAGCCGCGGCG 151
Oy      3749  GTGACCTCGGCGCTTAGGCGCCCGACCAACGCGCGCGGCGCGGTGCGCTCGGTCGTGGCCGCG 3808
Db      150  GCGAGGCTGTGGCCCGACCGCGGTGAGACCGCGCGCCCGCAATGGGCTCCGGGGGTGGCGCGAG 91
Oy      3809  TCCGCGATCTCCGTGCGTACCGCGGCTCGCCGTAGAGTGTCTTATGAC 3859
Db      90  CGTTCCGCGCGGGGCGGTGGCGGAGAGGCGCGGGGCGGTATTCGAGAGCGTC 40

RESULT 12
US-10-160-539-24/c
; Sequence 24, Application US/10160539
; Publication No. US20030162262A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120

```

## RESULT 12

US-10-160-539-24/c

Sequence 24, Application US/10160535

sequence 24, APPLICATION US/101

Publication No. US20

GENERAL INFORMATION:

GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary

APPLICANT: ASHLBI, Gary  
APPLICANT: BETTACH, Melanie C

APPLICANT: BETLACH

APPLICANT: BETLACH, Mary C.

APPLICANT: BETLACH, Mary C.



PRIOR FILING DATE: 1998-05-28  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 1565  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-10-468-24

Query Match 8.2%; Score 729.4; DB 8; Length 1565;  
Best Local Similarity 68.1%; Pred. No. 1e-153;  
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

2369 GGGGCGGTGACAGCGCGTTTCAAGACCGCGCGATCAAGCGCCCGATGCTGATCCACCCTCG 2428  
1530 GGGGGGATTCGGGGGGGGGGGGGGTCAAGCGACGAGAGCGGGGCTCTCCACGACGGTCC 1471  
2429 GCGACGTCCGCTTCAAGTTGTTGAGCGCGGCGGTCAACCGACTGTGTAACCGTCAAG 2488  
1470 GCGGCGTCCGCTCAAGCTGGTTCAAGCGGGCGGTGACGACTGATGAAAGCGCTCATG 1411  
2489 AAGAACTGTCCTCCCGGCTGCGATTCGATCTGCGGCGCGACGTCGAGAAAGTCTGAGCG 2548  
1410 AAGTACTGTCGCGTCAAGCGCGCGCACTGCGCGCGCTGACGAAAGTCTGAGCG 1351  
2549 ACCGAGTCAAGCTCCGCTCCGCGGTGACCGCGCGCGACGTAAGCGGTTGACCGCGCG 2608  
1350 ACCCTCGTGAAGAGAGTGTCCGGGGGTCAAGCGCGCGCGAGTGAAGCGGGTCCGCGCT 1291  
2609 AGCCCGGAGAAACCGGCGCTCCCGGTCAAGTGAAGTCTCCGAGAGGTGATTTGAC 2668  
1290 AGGTCCGGGAAAGCGCGCTCCCGGTACAGTACAGTCCGAGAGATCACTGAC 1231  
2669 GCCACTGCGGATGAGGCGGTGAGGCGGATCGTTTCCGCGCGGATCCGCAACGCTGAGCG 2728  
1230 GCGACTGCGGAGTGGCGGTGAGGCGGATCGTGAAGTGGCGGCTTGAATCCGAGAGTTCG 1171  
2729 TCAGACCCCGACGAGCTCTCCAGCGCTGAACCGAGTCTGCTGATCCCGGAGTTC 2788  
1170 TCGGCGCGCGTCCGAGCTCTTCAAGCGGTGAAGCGGTGAATGTGAGTCCGGGGGT 1111  
2789 CGCTCGCGGAGTGAATGACGACCGGCGAGGCCCTCCCGCAACTCGCTTCGCGCC 2848  
1110 CGCTCGCGGACCCCGCTCTCGAAGCGGTGAAGCGCTCTGAGACTCGCGCTCCCTCC 1051  
2849 TCGACAACTGCGCGTCTCCGACCGCTGATGCTCTCGCGAGCGGTGAACGAGTCAAGC 2908  
1050 TCGGCGAGTTCGCGTCTGACGCGCGCTGATGCTCTCGCAATGTTGACGAGTCAATC 991  
2909 GCGCGGTGCGGGTGAAGTCTGTTGAGTTCGCGAGTGAAGTGAACGAGTCAAGTCAAGCG 2968  
990 GTCTGCGCTGCGCGCGCTCTTGAAGTTCGCGAGTGAAGTGAACGAGTCAAGCGAGCG 931  
2969 TTGCGCGCGCGCGCGAGATGATGTTGAGCGCGAGCGCGACCGAGTCCGCGCGCTTCG 3028  
930 GAGGACGCGCGCGCGGAGAGATGATGAGCGGAGCGAGGTTGATCGCGGACTCCGCGCTCG 871  
3029 GCGCGCATTCGCGAGAGTCCCGAGAGTTCCTTCTGAGCCGTTGAGAGCGCGCTGAGTTC 3088  
870 GCGCGAGCTCTGGAAGCGCGCGGCTTCGCGGAGCGCGGAGAGCGCGCTTCCTTCG 811  
3089 CCGGAGTGGCGCTGATCTGCTGCTGTTGAGCGCGCTGACGAGAGTGAACGAGCGCGCG 3148  
810 CCGGAGTGTGCTGATCTCTCTGCTGTTGAGCGCGCTGAGCGAGTTCGAGTGGCGCTGC 751  
3149 AGGTCCCAAGAGCGCGCTGCGCGCGAGCTCTGTTGAGTGAAGCGCGCGCTTGTG 3208  
750 AGGCGCGAGAGCGCGCGCTGCGCGCTCAAGGAGTCTGCGTGAAGCGCGAGATTCG 691  
3209 TAGACGTGACAGGAGAGCGCGCGCGCGCGCTGCGAGCAACAGCGCGCGCTTACCGCG 3268  
690 TAGACGTGAGCGCGAGCGCGCTGCTGCTGCTGCGCGCGAGCTCCGAGCGCGCGCG 631

3269 TTGATCAACGTTTCCAGCGCGCGGAGAGTGAATGCGCGCTTCGAGAGCGTACC 3328  
630 TTGATGAGCGGCTCCAGCGCGCGGAGAGTGAATGATGATGATGATGATGATGATGATG 571  
3329 TCGTCAATGACCGAGGAGTGAAGTCTGCTTTCGCGGAGCGCGAGTCCGCGCTGATTCG 3388  
570 TCGTCAATGACCGAGGAGTGAAGTCTGCTTTCGCGGAGCGCGAGTCCGCGCTGATTCG 511  
3389 GCACTGTCAACCGGTAACGAGAGTGAACGAGAGTGAACGAGTGAACGAGTGAACGAGTGA 3448  
510 GCGCGGTGACAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 451  
3449 CCGACGCTGTAACGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3508  
450 CCGACGCTGTAACGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 391  
3509 TTGAGCGGCTGATGCTGCTTTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3568  
390 TCGAGCGGAGAGAGAGTGTCTTTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 331  
3569 CCGACGCTGTAACGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3628  
330 CCGAGCTCCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 271  
3629 TCGAACCGAGCGGCGATCTGCTTCAACGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAG 3688  
270 CCGAGGCTGATGAGGCGGCTCTTCAAGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 211  
3689 AGTTCCGCGAGCGGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3748  
210 AGCGGAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 151  
3749 GTGACTGCGCGCTGAGAGCGCGCGACCGCGAGCGCGCGCGCTGCGCGCTGCGCGCGCG 3808  
150 GCGAGGCTGCGCGCGCGAGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 91  
3809 TCCGAGTCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3859  
90 CGTTCGCGCGGCGGCGGCGGAGAGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 40

RESULT 14  
US-10-846-335-24/c  
Sequence 24, Application US/10846335  
Publication No. US20050233431A1  
GENERAL INFORMATION:  
APPLICANT: ASHLEY Gary  
APPLICANT: BERTLACH, Melanie C.  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 300622002121  
CURRENT APPLICATION NUMBER: US/10/846,335  
PRIOR FILING DATE: 2004-05-14  
PRIOR APPLICATION NUMBER: US/09/793,708  
PRIOR FILING DATE: 2001-02-22  
PRIOR APPLICATION NUMBER: US 09/657,440  
PRIOR FILING DATE: 2000-09-07  
PRIOR APPLICATION NUMBER: US 09/320,878  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: US 09/141,908  
PRIOR FILING DATE: 1998-08-28  
PRIOR APPLICATION NUMBER: US 09/073,538  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: US 08/846,247  
PRIOR FILING DATE: 1997-04-30  
PRIOR APPLICATION NUMBER: US 60/134,990  
PRIOR FILING DATE: 1999-05-20  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 1565









GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:28:59 ; Search time 1488.43 Seconds  
(without alignments)  
10776.903 Million cell updates/sec

Title: US-10-611-442-1  
Perfect score: 8942  
Sequence: 1 ggcgccttcacatccatcg.....gtgagcagccggagatcc 9024

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3032.6	33.9	8051	2	US-08-576-626A-2
2	2606.8	29.2	47981	3	US-09-679-279-1
3	1309.4	14.6	13613	3	US-09-105-537-3
4	729.4	8.2	1565	3	US-09-320-878-24
5	729.4	8.2	1565	3	US-09-141-908-22
6	729.4	8.2	1565	3	US-09-657-440-24
7	729.4	8.2	1565	3	US-09-793-708-24
8	722	8.1	1458	3	US-09-105-537-9
9	701.4	7.8	3756	2	US-08-576-626A-1
10	595.4	6.7	1248	3	US-09-105-537-7
11	519.2	5.8	5970	3	US-09-320-878-21
12	519.2	5.8	5970	3	US-09-141-908-11
13	519.2	5.8	5970	3	US-09-657-440-21
14	519.2	5.8	5970	3	US-09-793-708-21
15	461.2	5.2	1452	3	US-09-758-759-30
16	461.2	5.2	109519	3	US-09-758-759-1
17	397.4	4.4	80161	3	US-09-036-987A-1
18	397.4	4.4	80161	3	US-09-370-700-1
19	397.4	4.4	80161	3	US-09-603-207-1
20	381.6	4.3	6085	3	US-09-029-603-4
21	371	4.1	1518	2	US-08-660-765A-1
22	350.2	3.9	4496	3	US-08-765-907A-6
23	350.2	3.9	4496	3	US-09-987-614A-6
24	346.2	3.9	1194	3	US-08-765-907A-9

25	346.2	3.9	1194	3	US-09-987-614A-9	Sequence 9, Appl1
26	266.2	3.0	50937	3	US-09-428-517-1	Sequence 1, Appl1
27	255	2.9	11440	3	US-09-252-991A-2912	Sequence 2912, Ap
28	255	2.9	1425	3	US-09-252-991A-3053	Sequence 3053, Ap
29	238.2	2.7	38506	3	US-09-320-878-19	Sequence 19, Appl1
30	238.2	2.7	38506	3	US-09-141-908-1	Sequence 1, Appl1
31	238.2	2.7	38506	3	US-09-657-440-19	Sequence 19, Appl1
32	238.2	2.7	38506	3	US-09-793-708-19	Sequence 19, Appl1
33	212.6	2.4	1600	3	US-09-434-288-10	Sequence 10, Appl1
34	207.8	2.3	1281	3	US-09-105-537-19	Sequence 19, Appl1
35	196.4	2.2	8051	2	US-08-576-626A-2	Sequence 2, Appl1
36	192.8	2.2	12249	3	US-09-266-965-74	Sequence 74, Appl1
37	192.8	2.2	18331	3	US-09-266-965-96	Sequence 96, Appl1
38	189.6	2.1	44377	2	US-08-804-227C-7	Sequence 7, Appl1
39	189.6	2.1	44377	2	US-08-804-198-1	Sequence 1, Appl1
40	187	2.1	1224	3	US-09-266-965-22	Sequence 22, Appl1
41	183.2	2.0	1188	3	US-09-252-991A-936	Sequence 936, App
42	182.8	2.0	1332	2	US-08-660-765A-3	Sequence 3, Appl1
43	182.4	2.0	588	3	US-09-724-797-45	Sequence 45, Appl1
44	180.8	2.0	47981	3	US-09-679-279-1	Sequence 1, Appl1
45	178.6	2.0	1209	3	US-09-105-537-21	Sequence 21, Appl1

## ALIGNMENTS

RESULT 1  
US-08-576-626A-2/c  
Sequence 2, Application US/08576626A  
Patent No. 5998194  
GENERAL INFORMATION:  
APPLICANT: Summers, R.G.  
APPLICANT: Katz, U.  
APPLICANT: Donadio, S.  
APPLICANT: Staver, M.J.  
TITLE OF INVENTION: POLYMERIDE-ASSOCIATED SUGAR  
TITLE OF INVENTION: BIOSYNTHESIS GENES  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/576,626A  
FILING DATE: 21-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Diane Casuto  
REGISTRATION NUMBER: P-40,943  
REFERENCE/DOCKET NUMBER: 5857.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (847) 938-3137  
TELEFAX: (847) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8051 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-576-626A-2

[illegible]

Db	6484	CGCCGACATGATCATGCTTGGAGTTGTCCTGTGGCACCTGTGCATGATGCGGGCA	6425
Oy	3348	GGATCTGTTGCCGGGACCAACGACCTCCGCGTGTGACTTTGGACCTTGACCCGTGACG	3407
Db	6424	GCGTCTGTTGCCCGGGACCGGATGCGGCGCTCTGTAGCGGGACCGGTACCGGACGC	6365
Oy	3408	AGAGTGAACAGCGGAATGTCAGCTCGGGCGGGGTAGACCCGACCGCTGTACGGGAAG	3467
Db	6364	AGAGTGAACAGCGGAATGTCAGCTCGGGCGGGGTACAGGCCACCGCTTAGGGGAAG	6305
Oy	3468	CGGGCTCCCGCGGACCGGACCGCTCAACCGGCGGGCGGGTTCAGGGGCGCTGATCGTGT	3527
Db	6304	CAGGCTTGGGTAGACCGGCGCGTCCAGTGTGCTGTGGGGGTTCAGGGGCGCTTATCTGT	6245
Oy	3528	TGCTTCAGTACTTGGCCGCGGGGCGCTGTCCACCGCGGTGCCCCAGCTTGGGGATTCGAT	3587
Db	6244	TGACACAGTACTTGGCCGCGGGACCGTGTCTCACCGGGGTCCAGGTCGGGGATGCGGC	6185
Oy	3588	CGAACAGGTGAGCAGCTCCCGAACCCCGGCGCGGTTCAGGTGCAACCGAGGCACTCT	3647
Db	6184	CGAACAGGTGAGCAGGTGGCGCACTCCGGCGCGTCAAGGCCAGCGCGGCAACGCTCT	6125
Oy	3648	GCTCAACCGGGGTAAACGCGAGGCGCGTATGTGGGGGTGATGTTCGGCGAGCGGACCG	3707
Db	6124	GCTCAAGGGGTAAACGGGTGGCGCGCGGTATGTGGGGGTGATGTTCGGCGAGCGGACCG	6065
Oy	3708	CCTGTGTTCCGCGGTGTCCGATCCGACACCGGTGAGGGCGGTGACCTCCGCGCTGAGCG	3767
Db	6064	CTTCACGTGGGCTTCAACGTCACCGGCGCGGTGACCGGGAACTCCGAGCTCAGCG	6005
Oy	3768	CCCGCACACGACCGGCGCGGTCCGCGTGGGTGTGTCGGCGGTCCGCACTCCGTGGTA	3827
Db	6004	CCCGCACACGACCGGCGCGGTCCGCGTCCAGTCCGGGCGCGGTCCGCAACCCCGCTTCT	5945
Oy	3828	CGGGGTCCGCGGTAGAGTGTCTTCAATCCAGCGTGCAACCTTCGAGCGTGTGGTGC	3887
Db	5944	GGGGGTTCGCGTACGAGTGTGTCAATCGGGAGCGAACCTTCGACCTTGTGGGCGCA	5885
Oy	3888	GAGGATCAACGAACGTTTGCCTTCCGCTGTCCACTCCGCGTATCAACGACCGCTG	3947
Db	5884	GGTCTCAACGAACGTTTGCCTTCCGCTGTCTCTGTCCCATTTGGCGGTATCAAGTCCGCTG	5825
Oy	3948	GTGTGGAGAGTCGATGACGTTCGACACCCCGCGGATGCTCTCACTGAGACACCGAGGG	4007
Db	5824	GTGGGGGAGACCGGATGATGTCCGACACCCCGCGGATGCTCTGCTGTGACACCGCGGG	5765
Oy	4008	CCGTCGAGGAGGCGAGCACCCCTTCCGCGAGCCACTCCGCTCTGTGTACGCGGAGGGT	4067
Db	5764	CCGTCGAGGAGGCGAGCACCCCGTCCGAGGCGTTCGTGGTCTCCAGCGGACGCGC	5705
Oy	4068	GGCTCCGAGGATAGGGGTCACTGTGTGGAGCGGGGAGAGTAGAGGCTGTGACAG	4127
Db	5704	GGCTCCGAGGATAGGGGTCACTGTGTGGAGCACCGGGGAGAGTAGAGGTTGTGACCA	5645
Oy	4128	ACCTTCTCCGCGCGGAGATCGCTTCGACCGGTCACGTCGATCCGCTGCGGCTGCG	4187
Db	5644	ACCTTCTCCGCTCCGAGGATCGCTTCGACCACTGTGTCCGATGATGCGGATGCGCGAG	5585
Oy	4188	TCCACCAAGATGATACGCTACTTGGTATGTTGCTCTCTGTGGGCGGGAGCGATGACG	4247
Db	5584	TCCACCAAGATGATACGCTACTTGGTATGTTGCTCTCCAGGATCCAAAGCGGTGACG	5525
Oy	4248	GTGACCGCGGATACGTCCGCGAGCTCGCTGGTGTGAGCGGCGGTGTGACCCGGTTGTC	4307
Db	5524	GATATGCGCGGACGTCGCGGAGCTCGTCGAGTAGAGCGGTGTGATGAGCGGTGTGTC	5465
Oy	4308	TCCCTGTCTCCGCGCAACGCGTCGAGGAGGTGAGCCCATGATGCTCCGCGGCACTCTGCTC	4367
Db	5464	ACCTTGTCTCCGCGCAAGCGGTTCGAGGAGGCGCATTCGCGCGGCGCATCTGCTTC	5405
Oy	4368	ATCTTCCGTTGTGTCGATCTCGGTACCACTTGTTCGGGCGGATGCCGAATTTGTGC	4427

Db	5404	ATCTTGCCGGTGGTGCACCAATCCGATCAACAAGCTTGTCCGGTGCATCCCAAGTTGTGC	534
QY	4428	ATGGCCCGGATTCCTTTCCGGCCAGCAGCCCGCTGTCCGTGACCAACCGCCCGCTTGAAG	4487
Db	5344	ATGGCGCGGATCCGGTCCGGCCAGCAGCCCGCTGTCCGTGACAGATGGCCCGCCCTCGAAC	5285
QY	4488	GCGGGAACCGCCCTTGGTGGCCGGTGAAGCTGAACAACCTGGGCTCAACCGATCCGCCAAC	4547
Db	5284	GAGGTGAACCGCTTGTGTGGCCGTGGAAGCTGAACACTTCCGGCTTGCCAAGCCGCACAC	5225
QY	4548	GGACGTCACACCCGTGTGCAGCCCAAGGCGTGGGCGGCTCGAAGAAAGCTTGACCTGG	4607
Db	5224	GGCCGCCCGCCGGCCGGTCAAGCCAGCCGCGTGCGGGCTCGAAGAAAGTTTGACCTGG	5165
QY	4608	TGCTCCGCGCGGATCTTCCGCAAGCCGCTTCAACAAGTGTGGTCCGCCCAACAAGTGTACG	4667
Db	5164	TGCTCCGCGCGGATCTTCCAGCCGCTTCAACCGAAGCGGGCTGTCCCAAGGGTGAAG	5105
QY	4668	CCGAGATACCGCCCGCGTCTGTCCGGGTGAACAGCGCGGGAAGTGGTCCGGGTCCACACGA	4727
Db	5104	CCGAGATACCGCCCGCTTCCGGGTGAACAGCCACACGTAAGTGTCCGGGTCCAGCAGG	5045
QY	4728	CCGGTCGCCGGGTCCAGCTGCAGAACACCGGTGAAGTCCGAGCCAGCTGCCCGCGAG	4787
Db	5044	CCGGTCTCCGGGTCACAGTCCAGAACACCGTTTCACGCCCCAGCCACACTGCCCGGTGC	4985
QY	4788	GCGGTAGCCCGGAAAGTCAATGACCGGACATGATCACTTTCACCGGTGACCTCACCGGCCG	4847
Db	4984	GCGGTAGCCCGGAAAGTCAATGAAAGGACATGAACACTGCCCGGACACCTCCCTCCGGCG	4925
QY	4848	AGCACCACTTCCAGGGCCGACCGGTGGCTTGACAGTGGCGATGCAAGTGCCTGACCCCAAC	4907
Db	4924	AGCACCACTTGCAGGCCACCGTCCGCTTGGAAGGTGGCCACGCAAGTGGCGGACACCCGCC	4865
QY	4908	AGGTCCGGGACCCCGGCGCTCCGAACCTCCCGGACCAAGGGGCGCGCGTTGGTAGACAGTTG	4967
Db	4864	AGGTCCGGGACCCCGGCGCTTCCGAACCTCCCGGACCAAGTGGTCCGCGTTGTGACCAAGTTG	4805
QY	4968	TTGTTTCAGGGGCCACTCCAGGCGGGCCAGAACCGCTGCCGGTGCCAGTCCGATCCGACCGG	5027
Db	4804	TTGTTTCAGGGGCCACTCCAGGGCGGGCCAGAACCGCTGCCGCTCCCGACCGTCCGCGCTG	4745
QY	5028	CCCAAGTGCAGGGGGGTGCAGGAAGCGCTTCGCGACCGCCGAAGATCCGATCCGATCCGAT	5087
Db	4744	CCCAAGTGAAGGGGTGTGCAGGAATGCTCCGGGGCCGCCGAAGATCCGCAAGTCCGATCCGAT	4685
QY	5088	AACGGGCTTCAAGCCGTCGCCCGGTTGTGAAGCGCGGACACGCCACCGACAGAGCTTGCGCA	5147
Db	4684	GCGCGCTTTCATCCCGCGGATCTCGGTGTGAAGCGCGGACCGACGATGATGCTTGCGCA	4625
QY	5148	GCTGATTTGACGATGATGAGCTGTGCAGGACAGTTCGGTATCTGGCCGAAGGATCACTC	5207
Db	4624	GCTGATTTGACGATGATGAGCTGTGCAGGACAGTGTATCTGGTCAAGGATCACTC	4565
QY	5208	ACCGGTGTCGAGGCGCACCGCGCTGTGAAAGTCCCGCGGACCTTCGACGAGCATGTACC	5267
Db	4564	ACCGGTGTCGAGGCGCGCACCTGGGCTGGAAGTCCGCGGACCTTCGATCAAGATGTACC	4505
QY	5268	GGTTCCTCGTTCCGATGAACCGGCCCACTCTCTGGAATGCACAGCGCTCTGATGCGATCT	5327
Db	4504	GGTTCCTCGTTCCGATGAACCGGCCCGCTCTCTGGAATGTCAATGCGCTCTGATGCGACT	4445
QY	5328	GGGTACGCGGACGTCGACGATGATGAGGTAGATCGGCGGGGTTCTCCGGGCACT	5387
Db	4444	GCGAGCGCGGACGTTCAACACGATGATCAATAGCCGAGTCGGAATCTCTCCGGCGGCT	4385
QY	5388	CGGTGTATTGTCCGGCTGACAGTGCACCGTCCGCCCACTCCGACGACGTCGTGGCCAC	5447
Db	4384	CGGCTAGATTGTCCGGCTGCGCAGTGCACCGTAGGGCGCACTCCGCGACGTCATTCGCC	4325
QY	5448	CCGCCTCGGTGCAGAGTGCACAGCGCTGGAAGGGTCCGTCAATCTCTTGACACGGA	5507
Db	4324	CCGCCTCGGTGCAGAGTGCACAGCGGTGCACAGTGCACCCTCAATGTCTTGACACGGA	4265

QY	5508	GGGAGGAGCATGCGGTGTGGCGGGGGAGGACAAGGTTGATCCAGAACCTTCACTCCG	5567
Db	4264	GGGCGAGAGGCGCGTTGTGGCGGGGCGAGGACAGCGGCTGATCAAGATTGACCTCCG	4205
QY	5568	GGTGACTGTGTGTCAACGAGACCGCCGAAGATGTCTGAAGTACTTCTTGTCTCTGTTCGA	5627
Db	4204	GGTGCTGTGTGTGCGCACCGGTGACCGCCGAAGAGTGAAGTACTTCTTCTCTGTGTCTGA	4145
QY	5628	TGCGGTCTGTGCGCGCGGATCCACCCGCTGCGGTGAGTGTGGCCCAAGGCGGGTGGCTT	5687
Db	4144	TGCGGTCTGTGCGCGCGGATCCACCCGCTGCGGTGAGTGTGGGACAGGCGGGCGCGCT	4085
QY	5688	GGAAGAACTCACTGTGTGGTGGCTACGTCCGAGATCCAGCTCAGCAGGGGTGTCAATGTTGT	5747
Db	4084	GCAAGAACTCCGCGCTGTGGCGGACGTCGAGATCCAGCTCAGCAGGCTGTCAATGTTGT	4025
QY	5748	GGAACCGGGGTGCCCGCCCGACGAGGAAACGACAGACTGGGTCTCGAAGACCCCTGGG	5807
Db	4024	TGGCGGGGTGCCCGCGCCGATGAAACAGTGCAGCACCGGGCTTGAACAGGCTTCGG	3965
QY	5808	GCAATCCGTCCAGTAAACCGCGCGACCGGCCCCGGTTCAACCTGTGTGGATGTGAGGCCA	5867
Db	3964	GCAAGGCGCGCAGACACTGTCGCGGCGCGGTCCGCTCAACGGCGGTGGAGACGAGGCCA	3905
QY	5868	GCAAGGACCGCAGGTTCATGTTTCACACAGTTGTGTGAGCGAGCATCGCCGACAGCTGGG	5927
Db	3904	GCAAGGTGCGCAGGTTCATGTTTCACACGTTGTGTGTGTGACAGATGCGCCGACGCTGG	3845
QY	5928	CGAGGGTGAACCACTTGCAATTTCCGGGTCTCGGCGGGGTCTCGAAGACCTCGACCA	5987
Db	3844	CGACGGTCAAGCAACGGAAGTTCCGGGTGCTCGGCGAGTGTCTGAAACACTCGACGACCA	3785
QY	5988	TGTTGGGTTGACGTTTGCGACAGAAACGGAACCTGCTCGAACTGGAGGACGTCGACGA	6047
Db	3784	TGTTTCGGTTGCGCTTGCGACAGAACCAACGCGCTGTCTGGACTGGACACTCGACGCA	3725
QY	6048	GGAATCCGGTGGGGCGGGGTCCCGTCGAAGTACTCGATGAACTGGAACGCGGAATCCGTTGT	6107
Db	3724	GGAATCCGGCTCGGGCGCGGTGCGCTGTGAAGTAACTCGAAGAACCGGACTTTCGAGCGCGGT	3665
QY	6108	GGAACCTCAATGATATTTGCTGCGGGTGGCGCGAGGGTCGGCGACGCTGGAACCGGCTTGA	6167
Db	3664	GGAACCGGTGTATTTGCTGTGGGGTCCGCTGCAAGGGTCGGGAGAGGTGGAACCGGCTTGA	3605
QY	6168	TGTTGCAAGTTTCGGGCTTGAGCTGCACACAGCGGTGCAATTAACGCTCGAATCCCGCA	6227
Db	3604	TGTTGCCCGGTGCGGCTTTGGCTTGGCGCGACAGCAAGTCACACACACTCGAATCTCTTGA	3545
QY	6228	CGATCAAGACGAGGAACCCGATCTCGGGTTGGAACGATGATGGTTGATCCAGTCCCGTC	6287
Db	3544	CGATGAGGCGGAGGAACCGATCTCGGGCTGCACAGTATGAGGCTGTGATCCAGTCCCGCC	3485
QY	6288	GCATTCGGAATTTGTTGCGGACGTCGAGGCGCTCGATGTGAGGAAGAAAGCGCGCGAGTCCGT	6347
Db	3484	GCCAGCGGAATTTGTTGCGGACGTCGAGGCGCTCGATGTGAGGAAGAAAGCGCGCGAGTCCGT	3425
QY	6348	GCGGCAGACCGACCGTCTCTCGGGTGAACGACAGCGTTTCAATGTGTCTGAAGGACATC	6407
Db	3424	GCGGCAGACCTTGCGGTCTCTCGGGCTGGAACGACAGCGGTTCATGTGTCTGAACGGAACGC	3365
QY	6408	GCTTCACCTGGAAGCCGATTTGCTTCGGCGCGCGGTGGGCGCAACAGTCGTGACAGTCCGTCCG	6467
Db	3364	GCGGCACCTGGAAGCGCGTCTTCGGCGCGCGGTGGGCGGCAACCGAGTGTGTGTTCGCGCG	3305
QY	6468	TGGCGGTGGAAGGTTCGCGCTGCGGTCGAGTCAAGAAACGTTATTTGCGGATTTGCTGCG	6527
Db	3304	TGGCGGTGGAAGATTTCCGCGGTGCGGTCGAGGAAACGCGTTTGGCGGTCTGGAAGATTTCCG	3245
QY	6528	GAGTTCGATGACGTTTGAACCGGATCCCGCATACGCGCTCTCCGTGATGTGTGTGGGCGTCC	6587
Db	3244	GAAATGTCGCGGTGCGCGGTCCGATCAACCATTCGACTTCTCTTCGCGGTGTCGTGACGCGCT	3185



LOCATION: (33052)...(33112)  
 OTHER INFORMATION: megAII, ACP4  
 NAME/KEY: CDS  
 LOCATION: (33666)...(43271)  
 OTHER INFORMATION: megAII; SEQ ID NO: 15= translated amino acid sequence  
 NAME/KEY: misc feature  
 LOCATION: (33780)...(35027)  
 OTHER INFORMATION: megAIII, KS5  
 NAME/KEY: misc feature  
 LOCATION: (35385)...(36419)  
 OTHER INFORMATION: megAIII, ATR5  
 NAME/KEY: misc feature  
 LOCATION: (37068)...(37604)  
 OTHER INFORMATION: megAIII, KRS  
 NAME/KEY: misc feature  
 LOCATION: (37860)...(38120)  
 OTHER INFORMATION: megAIII, ACP5  
 NAME/KEY: misc feature  
 LOCATION: (38187)...(39470)  
 OTHER INFORMATION: megAIII, KS6  
 NAME/KEY: misc feature  
 LOCATION: (39795)...(40811)  
 OTHER INFORMATION: megAIII, ATR6  
 NAME/KEY: misc feature  
 LOCATION: (41406)...(41936)  
 OTHER INFORMATION: megAIII, KR6  
 NAME/KEY: misc feature  
 LOCATION: (42168)...(42425)  
 OTHER INFORMATION: megAIII, ACP6  
 NAME/KEY: misc feature  
 LOCATION: (42585)...(43271)  
 OTHER INFORMATION: megAIII, TR  
 NAME/KEY: CDS  
 LOCATION: (43268)...(44344)  
 OTHER INFORMATION: megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;  
 OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence  
 NAME/KEY: CDS  
 LOCATION: (44355)...(45623)  
 OTHER INFORMATION: megCIII, desosamine glycosyltransferase;  
 OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence  
 NAME/KEY: CDS  
 LOCATION: (45620)...(46591)  
 OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,  
 OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;  
 OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence  
 NAME/KEY: CDS  
 LOCATION: (46660)...(47403)  
 OTHER INFORMATION: megH, TEII; SEQ ID NO: 19= translated amino acid sequence  
 NAME/KEY: CDS  
 LOCATION: (47411)...(47980)  
 OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence  
 US-09-679-279-1

Query Match 29.2%; Score 2606.8; DB 3; Length 47981;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6415 CTGAGCCGATGCTGGCGGCGCGGTGGCGCAACAGTCGTGACCTGTCGGTGGCGGT 6474  
 DB 1 CTCGACCCGATGCTCGCGCGCGCGGTGGCGCAACAGTCGTGACCTGTCGGTGGCGGT 60  
 QY 6475 GGGAGGTGCGCGGTGCGAGTCAGGAAGTATGCGATTGCTGATGATTCGCGAGATCC 6534  
 DB 61 GGGAGGTGCGCGGTGCGAGTCAGGAAGTATGCGATTGCTGATGATTCGCGAGATCC 120  
 QY 6535 ATGACCGTTGACCCGATCCCATAGCCCTCTCCGATGATGTCGGGCGGTCCGTGCG 6594  
 DB 121 ATGACCGTTGACCCGATCCCATAGCCCTCTCCGATGATGTCGGGCGGTCCGTGCG 180  
 QY 6595 TACCGCCGGAATGATGATTCGATCAAGACCCCGCCAGTGTAGGGCTCCGCGCGCA 6654  
 DB 181 TACCGCCGGAATGATGATTCGATCAAGACCCCGCCAGTGTAGGGCTCCGCGCGCA 240

QY 6655 CGGAGAGAGTCCGTCGAACAATTCCGGGTGACCGGTCGCGGCGGTGAAACGGGC 6714  
 DB 241 CGGAGAGAGTCCGTCGAACAATTCCGGGTGACCGGTCGCGGCGGTGAAACGGGC 300  
 QY 6715 GTGAGACACCCGATCATTCGTGCGGTGAATCTTCACTGTGCGGCGCAATCTTTC 6774  
 DB 301 GTGAGACACCCGATCATTCGTGCGGTGAATCTTCTTCACTGTGCGGCGCAATCTTTC 360  
 QY 6775 TGAACCGGTGTTCCGTTGATGACCGGTTCCCGGCGGTCGGAATCTGCGTGGACT 6834  
 DB 361 TGAACCGGTGTTCCGTTGATGACCGGTTCCCGGCGGTCGGAATCTGCGTGGACT 420  
 QY 6835 GACCGGTGCGGCGTGTTCGCGCGGTTCCGAATGCGGATTCGTGATCGGCAAGTG 6894  
 DB 421 GACCGGTGCGGCGTGTTCGCGCGGTTCCGAATGCGGATTCGTGATCGGCAAGTG 480  
 QY 6895 GAGCGGTTGCTGACCGGATGATGCAATGAGCGCTCAATGCAATCTTGTAG 6954  
 DB 481 GAGCGGTTGCTGACCGGATGATGCAATGAGCGCTCAATGCAATCTTGTAG 540  
 QY 6955 CATGTCGCGCGCGGATCCGACGCGCGGAAACCGCGCATCCAGCTTTCGACGA 7014  
 DB 541 CATGTCGCGCGCGGATCCGACGCGCGGAAACCGCGCATCCAGCTTTCGACGA 600  
 QY 7015 CGTGCATCATCCGTCGAAGCGCGGATGACACCGACCAACGCGCATGTCGCGCACT 7074  
 DB 601 CGTGCATCATCCGTCGAAGCGCGGATGACACCGACCAACGCGCATGTCGCGCACT 660  
 QY 7075 GGAAGGTGCGCGATCAAGGAATGCGGTGTCATGACAGACCGCAACAGCTTTC 7134  
 DB 661 GGAAGGTGCGCGATCAAGGAATGCGGTGTCATGACAGACCGCAACAGCTTTC 720  
 QY 7135 GGGCTGCGGAAACAGATGATGTCGTCAGCGGTCATTGCCCCGCGCACCGCTT 7194  
 DB 721 GGGCTGCGGAAACAGATGATGTCGTCAGCGGTCATTGCCCCGCGCACCGCTT 780  
 QY 7195 GGAATCCGTCGACCGGTCGTCGAGTGCAGATGCGGGAACCGGTTTCGAGACGCA 7254  
 DB 781 GGAATCCGTCGACCGGTCGTCGAGTGCAGATGCGGGAACCGGTTTCGAGACGCA 840  
 QY 7255 GGTAGTAGCGATGACGAGGCTTCTGTCGCGCGGACGCGTGCATAGAGTGAATCCG 7314  
 DB 841 GGTAGTAGCGATGACGAGGCTTCTGTCGCGCGGACGCGTGCATAGAGTGAATCCG 900  
 QY 7315 TCACAGTCTTCAATCCGGAAGGCTTCTATGACAGTTGGCGATGAAAGCGCGCG 7374  
 DB 901 TCACAGTCTTCAATCCGGAAGGCTTCTATGACAGTTGGCGATGAAAGCGCGCG 960  
 QY 7375 GAGTTCAGATGCGCGGCTTCTATGAGGCTTGGTGCCACCGGATCTGTACTCG 7434  
 DB 961 GAGTTCAGATGCGCGGCTTCTATGAGGCTTGGTGCCACCGGATCTGTACTCG 1020  
 QY 7435 ATGCTCTGTCGGAAGGACGACGACCCCTGACCTGATGCAAGGTTTCGCGCGCGC 7494  
 DB 1021 ATGCTCTGTCGGAAGGACGACGACCCCTGACCTGATGCAAGGTTTCGCGCGCGC 1080  
 QY 7495 GGAACGGAACCGTACGCGATGCGGCGGAAACGTCGGTGTGATGCAACCGGACGCG 7554  
 DB 1081 GGAACGGAACCGTACGCGATGCGGCGGAAACGTCGGTGTGATGCAACCGGACGCG 1140  
 QY 7555 GCGAGGTGCTGCGGATCCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 7614  
 DB 1141 GCGAGGTGCTGCGGATCCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 1200  
 QY 7615 CAGTGGCCCACTGCGCGGCTTCTGAGCGCGCGCTTCCGAGATTTCAAGCGCGC 7674  
 DB 1201 CAGTGGCCCACTGCGCGGCTTCTGAGCGCGCGCTTCCGAGATTTCAAGCGCGC 1260  
 QY 7675 ACCGAGAACGCGCGCTGCTGATGACGTCGACGCGGATTCAGACGCTGTCGCGCA 7734  
 DB 1261 ACCGAGAACGCGCGCTGCTGATGACGTCGACGCGGATTCAGACGCTGTCGCGCA 1320  
 QY 7735 CTGATGACCGAGCTGCGGCTTCTGATCTGAGGAAGACTTCCCGGAGAGTCCCG 7794



Db 1321 CTGGTGAACGAGCTGGGGTGGCGCTGATCTGTGAACGACTTGGCCGGAGGATCCG 1380  
Qy 7795 GTGTGGGGGCTGGGTACCGGCGCGGCACTGAAGGGGCTGGACCCCGACCGCTCCGGTCC 7854  
Db 1381 GTGTGGGGGCTGGGTACCGGCGCGGCACTGAAGGGGCTGGACCCCGACCGCTCCGGTCC 1440  
Qy 7855 TGAACCTCGGCGACCGCGGTATGCTGAGCGCCAGGTCAAGCCCGCAACAGCTCGCGTG 7914  
Db 1441 TGAACCTCGGCGACCGCGGTATGCTGAGCGCCAGGTCAAGCCCGCAACAGCTCGCGTG 1500  
Qy 7915 ACCGAACAAGGCGCTGACCGCCCTGAGAGATGACCGCGGTCAACCGCGGTGCGGAAGCC 7974  
Db 1501 ACCGAACAAGGCGCTGACCGCCCTGAGAGATGACCGCGGTCAACCGCGGTGCGGAAGCC 1560  
Qy 7975 GCGGTGCTGGTGGGGGGTGGTGGCGGAGTGGCGGCGCAACCGGGGGCAACCGCGTCTG 8034  
Db 1561 GCGGTGCTGGTGGGGGGTGGTGGCGGAGTGGCGGCGCAACCGGGGGCAACCGCGTCTG 1620  
Qy 8035 GCCGTCAACCGAGCTTCCGAACTGGCGGCAAGCTTCCGAGCGACCGCGAGACCGCGAC 8094  
Db 1621 GCCGTCAACCGAGCTTCCGAACTGGCGGCAAGCTTCCGAGCGACCGCGAGACCGCGAC 1680  
Qy 8095 CGTGTGTGAGCGAGGTGTGCGGAGCGAGTCCCGGCGTCACTTGAACCGCGACCGCC 8154  
Db 1681 CGTGTGTGAGCGAGGTGTGCGGAGCGAGTCCCGGCGTCACTTGAACCGCGACCGCC 1740  
Qy 8155 GCGTTCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8214  
Db 1741 GCGTTCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800  
Qy 8215 GTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8274  
Db 1801 GTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860  
Qy 8275 CGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8334  
Db 1861 CGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920  
Qy 8335 GCGCTGTGTGCGACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8394  
Db 1921 GCGCTGTGTGCGACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980  
Qy 8395 CTGTCCCGGCTTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8454  
Db 1981 CTGTCCCGGCTTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040  
Qy 8455 CGTTGCCCGGTGAGCTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8514  
Db 2041 CGTTGCCCGGTGAGCTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
Qy 8515 CAACAGCATCTGTTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8574  
Db 2101 CAACAGCATCTGTTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160  
Qy 8575 GGTACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8634  
Db 2161 GGTACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220  
Qy 8635 CGTGCCTCGGTGAGCGAGTGAACCTTGTGAGGTGAGCGGCGGCGGCGGCGGCGGCGGCGG 8694  
Db 2221 CGTGCCTCGGTGAGCGAGTGAACCTTGTGAGGTGAGCGGCGGCGGCGGCGGCGGCGGCGG 2280  
Qy 8695 CGTGAAGTACATGCGGAGCCCTTGAAGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8754  
Db 2281 CGTGAAGTACATGCGGAGCCCTTGAAGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2340  
Qy 8755 GGAACCTCGGGGAGATGAGAGCAACCTTCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8814  
Db 2341 GGAACCTCGGGGAGATGAGAGCAACCTTCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2400  
Qy 8815 CTGCGTCAATGAGCGGAGTGTGAGATTCTGCGGCTCTGGCGGCGGCGGCGGCGGCGGCGG 8874

Db 2401 CTGCTCATGACGCGGATGAGTGTGCTGCGGCTCTGGCGGCTCCGATCTGATGCTGTG 2460  
Qy 8875 GGAAGCGGTGACCTTTCGCGCGCGCGGATGCGGCGCGGCGGATGACGGAACCCCGGACGCGG 8934  
Db 2461 GGAAGCGGTGACCTTTCGCGCGCGCGGATGCGGCGCGGCGGATGACGGAACCCCGGACGCGG 2520  
Qy 8935 GATGCTGTGGGGTTCGGAAGTCTGCGCACCGGCGCGGCGGCGGAGAGCTTCTTGGAGCTGTGCG 8994  
Db 2521 GATGCTGTGGGGTTCGGAAGTCTGCGCACCGGCGCGGCGGCGGAGAGCTTCTTGGAGCTGTGCG 2580  
Qy 8995 CCAACGAGAGTGGAGCAACCGGAGGATCC 9024  
Db 2581 CCAACGAGAGTGGAGCAACCGGAGGATCC 2610

RESULT 3  
US-09-105-537-3  
Sequence 3, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 13613  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-3

Query Match 14.6%; Score 1309.4; DB 3; Length 13613;  
Best Local Similarity 67.7%; Pred. No. 7.2e-206;  
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

Qy 2369 GGGGCGGTGACCGGCTTTCAGACGCGCGGATGAGCGCGCGGATGATCCACCGGTG 2428  
Db 10159 GGGGCGGTGACCGGCGGCTTTCAGACGCGCGGATGAGCGCGCGGATGATCCACCGGTG 10218  
Qy 2429 GCGAGTCCCGGTTGAGTGGTTGAGCGCGGCGGATGAGCGCGGATGATCCACCGGTG 2488  
Db 10219 GCGAGTCCCGGTTGAGTGGTTGAGCGCGGCGGATGAGCGCGGATGATCCACCGGTG 10278  
Qy 2489 AAGAACTGTCGCCCGGCTGCGGATGATGCTGCGGCGCGGATGATGAGCGGATGATGAGCG 2548  
Db 10279 AAGAACTGTCGCCCGGCTGCGGATGATGCTGCGGCGCGGATGATGAGCGGATGATGAGCG 10338  
Qy 2549 ACCGAGTGCAGGCTCCGCTCGGCGGATGAGCGCGGCGGATGAGCGCGGATGAGCGCGGATGAGCG 2608  
Db 10339 ACCGAGTGCAGGCTCCGCTCGGCGGATGAGCGCGGCGGATGAGCGCGGATGAGCGCGGATGAGCG 10398  
Qy 2609 AGCCCGGCGGAAACCGGCGCTCCCGGATGAGGATGAGCTTCCGAGCAGGATGATGAGCG 2668  
Db 10399 AGCCCGGCGGAAACCGGCGCTCCCGGATGAGGATGAGCTTCCGAGCAGGATGATGAGCG 10458  
Qy 2669 GCAACTGCGGATGAGGCGGCTGCGGCGATGATGCTTCCGCGGATGATGAGCGGATGAGCG 2728  
Db 10459 GCAACTGCGGATGAGGCGGCTGCGGCGATGATGCTTCCGCGGATGATGAGCGGATGAGCG 10518  
Qy 2729 TCGAACCCCGGAGCGGAGCTTCCAGCGGATGAGCGGATGAGCTTCCGAGCAGGATGAGCG 2788  
Db 10519 TCGAACCCCGGAGCGGAGCTTCCAGCGGATGAGCGGATGAGCTTCCGAGCAGGATGAGCG 10578  
Qy 2789 CGCTCGGCGGCTGATGAGAGACCGGCGGCGGCGGCTTCCGAGCAGGATGAGCGGATGAGCG 2848  
Db 10579 CGCTCGGCGGCGGCTTCTGAGAGCGGATGAGCGGCTTCTGAGCAGGATGAGCGGATGAGCG 10638  
Qy 2849 TCGAACCACTGCGGCTGCTCCGAGCGGCTGATGCTTCCGAGCAGGATGAGCGGATGAGCG 2908







```

US-09-141-908-22/c
; Sequence 22, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 30622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1565
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-22

Query Match      8.2%; Score 729.4; DB 3; Length 1565;
Best Local Similarity 68.1%; Pred. No. 7.5e-111;
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2369 GGGCGGTCGACGCGCTTCAGACGCGCGGATCAGCGCCCGGATGCTGATCCACCCGCTCG 2428
DB 1530 GGGGGGATCGGGGCGGGTGGGGGTGCGAGCGAGGAGCGCGGGCTCTCCACGCGCTTC 1471
QY 2429 GCGACGCTCCGCTTCGAGTGGTGAAGCGCGCGCTTCACGACTGCTGTAACCGTGCAGG 2488
DB 1470 GCGCGGTCGCGCTTCAGCTGTTCAAGCGCGGCGTGAAGCATGATCGAAGCGCTCATG 1411
QY 2489 AAGAACTCGTCCCGCGGCTCGGATGATGCTGCGCGCGGACGTAAGTGGTGAAG 2548
DB 1410 AAGTATCTCGTCCGCTTCAGAGCGCGCGCACTCGCGCGCGCTTCAGAGAGTCCCTGACG 1351
QY 2549 ACCGAGTCAGGCTCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 2608
DB 1350 ACTCGGTGAGGAGGAGTGGGGGTGCGGGGTGCGCGCGCGCGGTCGAGTGAAGCGGTCC 1291
QY 2609 AGCCCGGGGAAACCGGCTTCGCGGTACAGGTACGCTCTTCGAGAGAGTGCATCTGACC 2668
DB 1290 AGGTCCGGGAAAGCGGCTTCGCGGTACAGGTACGCTCTTCGAGAGAGTGCATCTGACC 1231
QY 2669 GCCACTGCGGATGAGGCGGTGCGGCGGATGCTTTCGCGCGGATTCGCGCAACACTGAGCG 2728
DB 1230 GCGACTGCGGATGAGGCGGTGCGGCGGATGCTTTCGCGCGGATTCGCGCAACACTGAGCG 1171
QY 2729 TCAGACACCCCGAGCGAGCTCTCCAAAGCGGTACCGAGTGGTTCGATGCGCGGAGTCC 2788
DB 1170 TCGGCGCGGTCGCGAGGCTGTTCAAGGCTTACCGTGAAGTGAAGTCCGCGGAGTCC 1111
QY 2789 CGCTCGCGGCTGATGTCAGCAACCGGCGAGGCGCTTCGCGCAACTCGCTGCGTCC 2848
DB 1110 CGCTCGCGGACCGGCTTCGAGAGGCGGTGAGGCGCTTCGAGAGCTCGCGCGCTCTCC 1051
QY 2849 TCGGACAACTGCGCTGTCGCGAGCGGTCGAGTCCGCGCGGAGCGGAGCGAGTGCAGC 2908
DB 1050 TCGGCGAGCTTGGCGTGTGTCAGCGCGGCTGATCTCTCGGAAATGTTGAAGAGTGCATC 991
QY 2909 GCGCGGTGCGGCTGAGCTCTGTTGAGTTCGCGCATGAAGTGCACAGGTCGATGAGCGCG 2968
DB 990 GTCTGCGCTGCGCGGCGTGTGAGGTGCGCATGAAGTGCACAGGTCGATGAGCGCGG 931

```

```

QY 2969 TTGCGCGCGCGCGGACGATGATGTTGAGCGCGGACCGGTCGCGCGCTTCG 3028
DB 930 GAGGCAACGCGCGCGGACGATGATGAGCGGACCGGTCGCGCGCTTCG 871
QY 3029 GCGCGATTCGAGAGATCCCGGAGTTCCTTCGACCGCTTCGAGAGCGCGCGCTTCG 3088
DB 870 GCGCGAGCTGTCGAGAGCGCGCGGATTCCTTCGAGAGCGCGCGCTTCG 811
QY 3089 CCGGTGTCGCTGATCTGCTGTTGTCAGCGCGGTCGAGAGGATGAGAGCGCGCG 3148
DB 810 CCGGTGTCGCTGATCTGCTGTTGTCAGCGCGGTCGAGAGGATGAGAGCGCGCG 751
QY 3149 AGGTTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3208
DB 750 AGGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691
QY 3209 TAGACGATGAGCGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268
DB 690 TAGACGATGAGCGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 631
QY 3269 TTGTCACGCTTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3328
DB 630 TTGTCACGCTTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571
QY 3329 TCGTTCATGACCGAGAGTATCTGCTTCGCGGACCGACCGACTCCGCTGCTT 3388
DB 570 TCGTTCATGACCGAGAGTATCTGCTTCGCGGACCGACCGACTCCGCTGCTT 511
QY 3389 GCACCTGTCACCGCTTCAGAGAGTGAAGCGGACGATGACGTCGCGCGCGCGGTTGAGC 3448
DB 510 GCGCGGTCACAGAGAGTGAAGCGGACGATGACGTCGCGCGCGCGGTTGAGG 451
QY 3449 CCGAGCTGTACGAGAGCGCGGCTCCCGCGTACCGGACCTTCACAGCGCGCGCG 3508
DB 450 CCGAGCTGTACGAGAGCGCGGCTCCCGCGTACCGGACCTTCACAGCGCGCGCG 391
QY 3509 TTGAGCGGCTGATCTGCTTCGAGTACTTCGCGCGCGCGCGCGCGCGCGCG 3568
DB 390 TTGAGCGGCTGATCTGCTTCGAGTACTTCGCGCGCGCGCGCGCGCGCGCG 331
QY 3569 CCGAGCTGTGAGAGTGAAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAG 3628
DB 330 CCGAGCTGTGAGAGTGAAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAGCAGTGAAG 271
QY 3629 TCAGACGAGCGGCTTCGCTTCGAGAGTGAAGCGGAGCGCGCTTCGAGAGTGAAG 3688
DB 270 TCAGACGAGCGGCTTCGCTTCGAGAGTGAAGCGGAGCGCGCTTCGAGAGTGAAG 211
QY 3689 AGTTCGCGAGCGCGGACCGGCTGCTTCGAGAGTGAAGCGGAGCGCGCTTCGAGAG 3748
DB 210 AGTTCGCGAGCGCGGACCGGCTGCTTCGAGAGTGAAGCGGAGCGCGCTTCGAGAG 151
QY 3749 GTGACCTGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3808
DB 150 GTGACCTGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
QY 3809 TCGCGATCTCGCTGCTGAGCGCGGTCGCGGTCGAGAGTGAAGTTCATTCAGC 3859
DB 90 CCGTTCGCGCGGCGGTCGAGAGGCGCGGCGGTCGATCGGAGCGCTC 40

```

```

RESULT 6
US-09-657-440-24/c
; Sequence 24, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

```

FILE REFERENCE: 300622002120  
CURRENT APPLICATION NUMBER: US/09/657,440  
CURRENT FILING DATE: 2000-09-07  
PRIOR APPLICATION NUMBER: 09/320,878  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
PRIOR FILING DATE: 1998-08-28  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 24  
LENGTH: 1565  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-657-440-24

Query Match 8.28; Score 729.4; DB 3; Length 1565;  
Best Local Similarity 68.18; Pred. No. 7.5e-111;  
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

2369 GGGGCGGTCGACCGCGCTTCAGACGCGCGGATCGAGCGCGCGGATCGGATCCGACCGGTCG 2428  
1530 GGGGGGATCGGGGCGGGTGCAGGCGGAGGCGCGGCGCTCCCTCCAGCGCTCC 1471  
2429 GCGAGTCGCGCTTCGAGTGGTTGAGCGCGGCGGTCACCGACTGTCGAAACCGTCGAGG 2488  
1470 GCGGCGTCGCGCTTCAGACTGTTGAGCGCGGCGGTCGAGCTGATCGAAGCGCTCAG 1411  
2489 AAGACTGTCGCGCGCGCTGCGGATCGATGCTGCGCGCGCGCGAGTCGAGTCGTCGAG 2548  
1410 AAGACTGTCGCGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1351  
2549 ACCGAGTCGAGCTTCGCGTTCGCGGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2608  
1350 ACCGAGTCGAGCTTCGCGTTCGCGGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1291  
2609 AGCGCGGAGAAACCGGCTCCCGGTACAGGTAGAGTCTCCGAGCGAGTCGATTCGACC 2668  
1290 AGGTCGCGGAGAACCGGCTCCCGGTACAGGTAGAGTCTCCGAGCGAGTCGATTCGACC 1231  
2669 GCGACTGCGGATGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTC 2728  
1230 GCGACTGCGGATGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTC 1171  
2729 TCGACACCGCGAGCGAGCTCTCCAGCGCGTACCGGAGTCGTCGAGTCGCGGCGGTC 2788  
1170 TCGGCGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTCGCGGCGGTC 1111  
2789 GCGTCGCGGTCGAGTCGAGCGAACCGGCGGAGCGCGCTCCCGCACTCGGTCGCTCGCC 2848  
1110 GCGTCGCGGAGCGCGCTCTCCAGAGCGGTCGAGGCGCTCTCGAGCTCGGCGCGCTCTCC 1051  
2849 TCGACACCTGCGCGCTCTCCAGCGGTCGAGTCTCTCGCGGAGCGGTCGAGTCTGAGC 2908  
1050 TCGGCGAGCTTGCCTGTCGACGCGCGCTGAGTCTCTCGCGAATGTTGACGAGTCTGATC 991  
2909 GCGCGGTCGCGGCTGAGCTCTGTTGAGTTCGCGGAGTTCGACCGAGTCTGATGAGGCG 2968  
990 GTCCTGCGCTTCGCGCGGCTGTTGAGTTCGCGGAGTTCGACCGAGTCTGAGCGAGGCG 931  
2969 TTGCGCGCGCGCGGAGGATGATGTTGAGGCGGAGCGGAGCGGAGCGGTCGCGCGCTTCG 3028  
930 GAGGCGAGCGCGCGGAGGAGCAGATGATGAGGAGCGGAGGTTGATCGGCGAGCTCGGCGTC 871  
3029 GCGCGCATTCGCGAGGAGTCCGAGGTTCTTCTGAGCGCGCTTCGAAAGCGCGACGCTTG 3088  
870 GCGCGCAGTGTCTGAGAGCGGCGAGTCTCGCGAGCGCGCGAGAGCGCGCTTCTTC 811  
3089 CCGGTCGCGCTGATCTGTCGTTGTTGACCGCGTACGAGGAGTACGAGCCGCGCG 3148  
810 CCGGTCGCTGCTGATCTCTCTGTCGTTGAGCGCGGTCGAGGAGTGTGAGTGTGTC 751  
3149 AGGTCGACGAGCGCGGCTGCGGCGGAGCGCTCTGTTGAGGCGGAGCGGTCGTCG 3208

750 AGGCCCCAGAGCGCGGCTGCGCTTCAGAGGTCGCTCGGTGAGCGCGGAGGTCGTCG 691  
3209 TAGACGATGAGGCGGAGCGCGCGCGCGCGCGCTGCGAGCAGAACCGCGGATACCGCG 3368  
690 TAGACGATGAGGCGGAGCGCGCGCGCGCGCGCTGCGAGGTCGCGGAGCGCGCGG 631  
3269 TTGTCGACCGCTTCGAGCGCGCGCGGAGGTCATGCGGTCGCGGTCCTTCGACGCGAC 3328  
630 TTGTCGAGCGGCTTCAGCGCGCGCGGAGGTCATGCGGTCGCGGAGGTCGCGGAGTC 571  
3329 TCGTCGATGACCGGAGGTCGTCGTCGCGGAGCGCGGAGCTCCGCGTCGTCGTCGTC 3388  
570 TCGTCGATGACCGGAGGTCGTCGTCGCGGAGCGCGGAGCTCCGCGTCGTCGTCGTC 511  
3389 GCACTGTCACCGGTCGCGGAGGTCGTCGTCGCGGAGCGCGGAGCTCCGCGTCGTCGTC 3448  
510 GCGCGGTCGACGAGCGCGGAGGTCGTCGTCGCGGAGCGCGGAGCTCCGCGTCGTCGTC 451  
3449 CCGAGCGCTGTCGAGGAGGCGGCGCTCCGCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3508  
450 CCGAGCGCTGTCGAGGAGGCGGCGCTCCGCGTCGTCGTCGTCGTCGTCGTCGTCGTC 391  
3509 TTGAGCGCGCTGATCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3568  
390 TCGAGCGGAGCGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 331  
3569 CCGAGCGCTGTCGAGGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3628  
330 CCGAGCTTCGCGGAGCTCCGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 271  
3629 TCGAACCGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3688  
270 CCGAGCTGTCGAGGCGCTCCGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 211  
3689 AGTTCGCGAGCGGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3748  
210 AGCGGAGCGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 151  
3749 GTGACCTGCGGCTGAGCGCGCGCGAGCGCGCGCGCGCGCGCTGCGGTCGTCGTCGTC 3808  
150 GCGAGGTCGTCGCGGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 91  
3809 TCGCGATCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3859  
90 GCTTCGCGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 40

RESULT 7  
US-09-793-708-24/c  
Sequence 24, Application US/09793708  
Patent No. 6902913  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETHLACH, Melanie C.  
APPLICANT: BETHLACH, Mary C.  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: RECOMBINANT MARBONOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 300622002121  
CURRENT APPLICATION NUMBER: US/09/793,708  
CURRENT FILING DATE: 2001-02-22  
PRIOR APPLICATION NUMBER: US 09/657,440  
PRIOR FILING DATE: 2000-09-07  
PRIOR APPLICATION NUMBER: US 09/320,878  
PRIOR FILING DATE: 1999-05-27  
PRIOR APPLICATION NUMBER: US 09/141,908  
PRIOR FILING DATE: 1998-08-28  
PRIOR APPLICATION NUMBER: US 09/073,538  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: US 08/846,247  
PRIOR FILING DATE: 1997-04-30  
PRIOR APPLICATION NUMBER: US 60/134,990  
PRIOR FILING DATE: 1999-05-20

```

; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1565
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-24

```

Query Match	8.2%	Score 729.4;	DB 3;	Length 1565;
Best Local Similarity	68.1%;	Pred. No. 7.5e-111;		
Matches 1015; Conservative	0;	Mismatches 476;	Indels 0;	Gaps 0

QY	2369	GGGGGGGTGCAAGCGCGTTCAAGACGGCGGGATTCAGAGCCCGAATGCTGATCCCAACCGGTGC	2428
Db	1530	GGGGGGGATTCGGGGGGGGGTGTCGGGTACAGCGCAGGAAGCCGGGGCTCTCTCCAGCGCTCC	1477
QY	2429	GCGACGTCCTCGTTGAGATTGGTTGAAGCGGGCGGTCAACGACGTGTGCGAAACCTGTGAGG	2486
Db	1470	GCGCGCTCGCGCTCCACGCTGGTTCAGGGGGGGCGGTGACGACCTGATTCGAAGCGTTCATG	1411
QY	2489	AAGAACTGTCCCCCGCGCTGGCGGATTCGATGCTGCGGCCCGACGTGACGAAATGCTGTGAGCG	2544
Db	1410	AAGTACTGTGTGCGCGGTGCAGCGCGCCGACCTGCGCGCGCGCTTCACGAAGTCCCTGACG	1351
QY	2549	ACCGAATGTCAGAGCTCCGGTCCGGGTGACCCGCGCGGACGATAGCGGGTGGCCCCCGCC	2606
Db	1350	ACCTTCGGTGAAGGAGATGTTCGGGGGTTCAGCGGCCCGCGGATGTAGCGGGTTCGCGCGCTCC	1291
QY	2609	AGCCCCGGGAAAACCGGCTCTCCCGGTACAGTGAAGCTCTCCGAGCAGGTTCATCTGACAC	2666
Db	1290	AGGTTCGGGGAAGCGGCTTCGCGGTACAGTACAGTTCGCGGAGGAGATTCGACTGACAC	1231
QY	2669	GCGACCTGGCGGATGGGGGGGTGCGGGGCAATCGTTTCCGGCCCGAATCCGGAACAGCTGGCG	2722
Db	1230	GCGACTGGGGGTGCGCGGTGAGCGCGCAATGTGTGGGGCTTGTATCCGACGATTCGGCG	1171
QY	2729	TCGACACCCCGACGACGAGCTCTCCACGCGGTAAACCAAGTGCAGTCTGCATGCCCGGGTTC	2788
Db	1170	TCGGCCCCGGTGGCGCAGGCTGTTCAAGGCGGTAGAGCGTATGTGATGTGAAGTCCGGGGGTG	1111
QY	2789	CGCTTCGGCGGCGTATGTGACGAACCGGGCGAGGCCCTCCGCAATCGGTCCGTTCCGCC	2848
Db	1110	CGCTTCGGGGAACCGGCTCTCCGAAGCGTGTGAGGGGCTCTCGAAGCTCGGCGCGCTCTCC	1051
QY	2849	TCGGAACAACCTGCGGTGTCGCGACCGGTGAAGTCTTCGCGGAGCGGTACGAAGTCCGAC	2908
Db	1050	TGCGGCAAGCTTGGCGGTGTCACGGCGCGCTGTAGTCTTCGCGCAATGTTCACGAAGTCCGATC	991
QY	2909	GCGCGGTGGCGGCTGGAATCGTTTGAAGTTCGCGGATGAGTGCACGAGTTCGATGAGCGG	2968
Db	990	GTCCTGCGCTGCGCGGGGTGTGAGGTTCGGCGATGAAATGTGACCAAGTTCAGCAGCGGG	931
QY	2969	TTGGCCCCCGCCCGGCAAGATGATGTGTTGAGGCCGAGCCGGAACCGGTGCGGCCCGTTCCG	3028
Db	930	GAGGCAACGGCCCGGGAGACACGATGTAGCGGAAGCCGAGTTGATTCGAGTCCGCGCTCCG	871
QY	3029	GCGGCGATTCGCGCAGGAAGTCCCGGAGGTTTCTTCGACCGGTTGGAAGGCGCCACGCTTG	3088
Db	870	GCGGCGATGCTGTGAAGCGGCGACGATTTCTCGCGAGCGCGGCGGAAGGCGGCTTCTTG	811
QY	3089	CCGGTGTGGCTGTGTAATCTGTCGTTGTTCAGCCCGTACAGAGAGTACGAGCCGCGCG	3148
Db	810	CCGGTGTGTGCTGTGTAATCTCTGTCGTTGAGGCGGTAGACGAGGTTCGATGGCGTGC	751
QY	3149	AGGTCCCAACAGCGCGGCTGCGCGGCGGACGTCGTTTCGTTGAAGGCGGAAGCGTTGGTG	3208
Db	750	AGGCCCCAAGAGCGCGGGCTGGCGCTTCAGAGGGTGGCTCGGTGAACCGCGCAAGAGATTCGTG	691
QY	3209	TAGACGGTTCAGGGGGAAGCCCGCGGGGCGGCTGTCACATCAACGCCCCGATACCGCGG	3268
Db	690	TAGACGGTGGCGCGCAAGCGCGTGTTCGGTGGCGTTCGCGGCGCAGAGCTCCAGGCGCGGG	631
QY	3269	TTGGTCAACGGTTCAGAGCGCGCGGAGAGATCAATCCGCTGGGGTCTTCGACGGCAC	3328

Db 630 TTGGTGAACGGCTCTCAAGCCCGCCGAGAAAGTACTCCGCAAGGGGTTTCCCGCGGGTATC 571

QY 3329 TCGTGCATGACCGAAGGTAGAGATCTCGTTCGCGGGCAACACCGACTCCGCGCTGCATCTTG 3388

Db 570 TGTGTGATGATGACCGAACCGGACATAGCGCTTTCGCGCGGTGAGAGGCGGAAGGGGTTCGTAAGCG 511

QY 3389 GCACCTTGTCACCCCGTACCGCAGAAATGACAGCGGAAATGCACTGTCGAGCCGAGGGTAAAGAC 3448

Db 510 GCGCGCGGTCAACACCGAACCGAAGAAATGGCAGCGGAAATGCAAGTTCGAGCCGAGGGTAAAGG 451

QY 3449 CCGACGCTGTACCGGAAAGCGAGGCTCCCGCGGTACCGCAGCGCTTCAACGCGCCGAGCGCG 3508

Db 450 CCGACGCTGTACCGGAAAGACGGGCTCTCTGCGGAGCGCGCGGTGAAAGACCGCCGCGCTGT 391

QY 3509 TTCAAGCGGCGCTGATTCGTGTGCTCTCAAGTCTTTCGCCCGCGGCGCTGTCTCAACCGCCGTG 3566

Db 390 TCGAGCGGAGACAGAGGTGTCTTCTCAAGTACCCCGCGCGAGGCGCGGTCTGCACCGCGGTG 331

QY 3569 CCCAGCTGCGGGAATGCGATTCGACAGGTTCAGAGCAAGTCCCGAACCCGAGCCCGCGTTCAGG 3628

Db 330 CGAGAGCTCCGAGACCTGCCCGAACAGAGCGGAGAGGCGCGCGAAGAGCGTCCCGGTTCAGCG 271

QY 3629 TCGAACCGACCGGCGCATCTGCTTCAACCGAGGTGAAACGAGCGAGCGCGCTAGTGTGGCGGTG 3688

Db 270 CCCAGAGTCAGTGGCGGGCTCTCTCAAGCGGGGTAAAGAGGAGCTGTGGCCGTACGCAACGCGCG 211

QY 3689 AGTTTCGAGGAGCCGGAACCGCGCTGTCTGTTTCGCGCGGTGTGGAATGCGGCAACCGGTGAGCGCG 3748

Db 210 AGCGGAGCAGAGGTATGGCGGGCGGTGCTTCCGCGCTCTGTTCGCGCGGACAAGAGCCCGCGGCG 151

QY 3749 GTGACCTTCGAGCGCTGAGCGCCCGGACCAACGAGCGCGAGCGGTGAGCGTTCGAGTCTGTGCGCG 3808

Db 150 GCGAGGGTCTTGGCCGAGCGGCGTGGACGCGCGCCCCAGATGGGCTTCGAGGGGTGCGCGAG 91

QY 3809 TTCGCGAATCTCCGTCTGCTGTAAGCGCGGTGCGCGGTAGAGATGGTCTTCAATCTGAC 3859

Db 90 CGTTTCGCGCGAGGCGGTGCGGAAAGGCGAGGCGCGGCGCGGTCAATCTGGAACGCTC 40

```

RESULT 8
US-09-105-537-9/C
Sequence 9, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and plitromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 9
LENGTH: 1458
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-9

```

Query Match	8.1%	Score 722	DB 3	Length 1456
Best Local Similarity	69.5%	Prod. No. 1.2e-109		
Matches 980	Conservative 0	Mismatches 430	Indels 0	Gaps 0
Qy	2395	GCGGATCAGGCCCCGATGCTGATCCACCCGCTGGGAGAGCTCCGTTGAGTTGTTGAG	2454	
Db	1455	GCGGAGGAGCCCGGAGCCTCTCTCCACGCCGTCGGCGGCTCGCGCTCCAGCTGGTTTCAG	1396	
Qy	2445	CCGGGCGGTACCCGACCTGCTCGAAGACCGTGAGGAGAGAACTCGTCCCGCCGGCTGGGATC	2514	
Db	1395	GCGGCGCGGTACGACCTGATCGAAGCCGTCACAGAGATCTCTCGTCGCGCTCGACGAGCCGC	1336	



QY 2515 GATGTCGCGGCGCGAGTGAAGTGTGAGCGACCGAGTGCAGGCTCCGCTCCGCGGCT 2574  
DB 1335 CACCTGCGCGCGCGCGCTCGAGGAAGTCTCTGAGACCTCGGTGAGGAGGTGTCGAGGAGT 1276  
QY 2575 GACCGCGCGCGCGAGTGAAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2634  
DB 1275 CAGCGCGCGCGCGAGTGAAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1216  
QY 2635 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 2694  
DB 1215 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 1156  
QY 2695 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 2754  
DB 1155 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 1096  
QY 2755 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 2814  
DB 1095 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 1036  
QY 2815 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 2874  
DB 1035 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 976  
QY 2875 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 2934  
DB 975 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 916  
QY 2935 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 2994  
DB 915 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 856  
QY 2995 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3054  
DB 855 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 796  
QY 3055 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3114  
DB 795 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 736  
QY 3115 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3174  
DB 735 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 676  
QY 3175 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3234  
DB 675 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 616  
QY 3235 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3294  
DB 615 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 556  
QY 3295 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3354  
DB 555 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 496  
QY 3355 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3414  
DB 495 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 436  
QY 3415 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3474  
DB 435 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 376  
QY 3475 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3534  
DB 375 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 316  
QY 3535 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3594  
DB 315 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 256  
QY 3595 CAGGTGAAGTCTCTCGAGCGAGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 3654

DB 255 GCGGAGGAGGCGCGGAGGCG 196  
QY 3655 CCGGAGGAGGCGCGGAGGCG 3714  
DB 195 CCGGAGGAGGCGCGGAGGCG 136  
QY 3715 CCGGAGGAGGCGCGGAGGCG 3774  
DB 135 CCGGAGGAGGCGCGGAGGCG 76  
QY 3775 CCGGAGGAGGCGCGGAGGCG 3804  
DB 75 CCGGAGGAGGCGCGGAGGCG 46

RESULT 9  
US-08-576-626A-1  
Sequence 1, Application US/08576626A  
Patent No. 5998194  
GENERAL INFORMATION:  
APPLICANT: Summers, R.G.  
APPLICANT: Katz, D.  
APPLICANT: Donadio, S.  
APPLICANT: Staver, M.J.  
TITLE OF INVENTION: POLYMERIDE-ASSOCIATED SUGAR  
TITLE OF INVENTION: BIOSYNTHESIS GENES  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Road  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/576,626A  
FILING DATE: 21-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dianne Casullo  
REGISTRATION NUMBER: P-40,943  
REFERENCE/DOCKET NUMBER: 5857, US, 01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (847) 938-3137  
TELEFAX: (847) 938-2623  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3756 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-576-626A-1

Query Match 7.8%; Score 701.4; DB 2; Length 3756;  
Best Local Similarity 66.0%; Pred. No. 3.2e-106;  
Matches 1118; Conservative 0; Mismatches 526; Indels 51; Gaps 5;

QY 7330 CCGGAGGAGGCGCGGAGGCG 7389  
DB 42 CCGGAGGAGGCGCGGAGGCG 101  
QY 7390 CCGGAGGAGGCGCGGAGGCG 7449

Db 102 CGGCGCTGCACTGGGGTTAAGGCAAGCAAGCCCTTACCCGATGCTGTGGCA 161  
 Qy 7450 CGGAGACGACACCCCTGACCTGTTAGAAAGTTGGGGGCGCGGAGACGGGTAC 7509  
 Db 162 CAGAGACGACACCCGACGCGCGGTACGCTCAATGCGAGTCCGGTGTGGCGC----- 216  
 Qy 7510 GCGAGTGGGCGGAGAGTGGGTGTGATGACCAAGCGGACCGCGCGGAGTGTGCGC 7569  
 Db 217 -----GAGAGACCGAGAGTGGGTGTGCGGACCAAGCGGACCGCGCGGAGTGTGAC 272  
 Qy 7570 GATTCGGGGCTTCAACCAAGCGCGCGCGGACGCTGCGCGGTGATGAGTGGCCACTGC 7629  
 Db 273 GACCCCGCGTTCACCCCGCGCACCGAGCAACCGGAAATGATGCGGCGCGCGCGCG 332  
 Qy 7630 CCGGCGGCGCTCTGAGGCGGCGCGCGCGCGGAGTTTCAAGCGCGCGCACCGAGAGCGCGCG 7689  
 Db 333 CACACCGCGCGAGTGGGCGCGAGCGGTTCGAGGAGCTTGACACCGCGGTCTTGGAGAG----- 387  
 Qy 7690 TCGGTGACATGAGACCGCGACCTGCTCAGCAGCGGTGCGCGAGGCTGTGACCGAGCTG 7749  
 Db 388 -GGAGGTTCCCGACGTTGGGGGAACTGGCGGAGAGCTTCCCGGTCTGCCGCGCGC 446  
 Qy 7750 GGGTCCGCGCTTCAATCTCTGAAAGACTTCCCGGAGAGTCCCGGTGTGGCGCTCGGT 7809  
 Db 447 GGGCGCGCGGTGAGACCTGCTGCGGAGCTTCCGCTGGGAGGTACCGGTGACAGGCGATGACC 506  
 Qy 7810 ACCGCGCGCGCACTCAAGGCGGTGAGACCCCGACCGTCTCCGCTCTGAGACTCGGCGACC 7869  
 Db 507 GCGGTCTCGGCGCGACGCGGAGTG-----CTGCGCGCGCGCGGTGGAGCGCC 554  
 Qy 7870 CGGATATGCTTGAACCGCCAGGTCAAGCCCGCAACAGCTCGCGGTGACCAAGACGCGCTG 7929  
 Db 555 CGGTCAAGCTTGAACCGCCAGCTCAAGCCCGCAAGAGTCCGGGTGACCAAGACGCGGTG 614  
 Qy 7930 ACCGCGCTCGACGATGACGCGGTGACCGCGGTGCGGAGCGCGCGGTGCTGTGGGG 7989  
 Db 615 GCGGCACTGCGCGCGCGCACCGCGCATGCGCGCC-----TGTTTC 653  
 Qy 7990 GTGGTGGCGGAGCTGGCGGCGCAACAGGTGGGCAACCGCTCTGTGCTGTACCGAGCTT 8049  
 Db 654 GCGGCGCGCGAGATGACCGGGAACACCGTGTGACGCGGTCTGTGCGGTCTGCGCGGAA 713  
 Qy 8050 CCGCAACTGGCGGCGACGACTTGGCGAGACCCCGAGACCGCGACCGCGTGTGACGAG 8109  
 Db 714 CCGGGGCTGGCGGACGATGCGCGAGACCCCGCGCGCGCGGACGAGACCGTGTGCGGAG 773  
 Qy 8110 GTGTCCGAGACGAGTCCCGCGCTGCACTTGAACCGCGCACCGCGCGGTGACCGCGG 8169  
 Db 774 GTGTGCGCGCTGACCGCGCATTTGACCTTGAAGCGGCGCACCGCGAGGTGCGG 833  
 Qy 8170 GTGGGGGGGTGCACTGCCGACCGGTGCGAGGTGACAGTGTCTGTCCCGCGGAGAC 8229  
 Db 834 CTGCGGAGACGATGCGGAGGCGAGAGGCGAGAGTGTGTCTGTGCGGCGGCGGAC 893  
 Qy 8230 CGTATCCGAGGCTTCAACCGATCCGACCGGTTGAGAGTGTGACCGTGGCGCGGAGCGC 8289  
 Db 894 CGGACCGGAGGCTTTCGCGAGCGCGACCGCTTGAAGTGTGACCGCGCGGAGCGCGAC 953  
 Qy 8290 GAGATCTGTGCTCCGCGCGCGCTGCGCGCGGACCGACTCGACCGCTGTGTGGCACCC 8349  
 Db 954 CGGCGCGCTGTGCGGACATGCGCGCGCACCGCGGCA-----GGTGAAGAGAGCTGTCAACCGG 1010  
 Qy 8350 CTGGGCGACGCGCGCGCTGGGGGCGCGCGCGCGGTGTGCTCCCGGTGTCCGTTCCGG 8409  
 Db 1011 CTTCGACACGCGCGCATGCGGCGCGCGCGGACGAGCGCTGCGCGGACCTCAAGCGCGG 1070  
 Qy 8410 CCGGTATGAGACGAGTGTGACCGCGGTGCGCGGTGTCTCAAGCGGTTGCGCGGTGAG 8469  
 Db 1071 CCGGTGTCTCGGCGCGCGCGCATACCGCTCTGCGGAGAACCAACGCTGCCCGGTGAG 1130  
 Qy 8470 CTGTAGAGAGAGATGCGCGGTGTGTTTTCATGATGAGCTGTCAACAGCCANTGTT 8529  
 Db 1131 CTCTGAGGATTCGCGGATGCGCGGTGTCTCTCTCATGAGCAGGAGAGGCACTCTT 1190

Qy 8530 CGGCTGTGCTCGCTGCAAGGCGCTTCCAGGCGCGCGGACAGAGGTAGGGTGTGCGC 8589  
 Db 1191 CGGCTGTGCTCGCTGCAAGGCGCTTCCAGGCGCGGACAGAGGTAGGGTGTGCGC 1250  
 Qy 8590 CTGCGGCGCGCTGACGACGAGCTCAACCGGTGCGGTCTTCAAGCGCGCTGTGAGTGA 8649  
 Db 1251 GTCCCGGCGCTCAACGAGAGCATCAACCGGCGCGGCTGTACCGCGTCCCGGTGCGCAC 1310  
 Qy 8650 CGAGCTGGAACCTTGTGAGTGGACCGCGCACCGCGGCGCAGAGCATGTGTGATGATGCG 8709  
 Db 1311 CGAGCTGGAACCTTGTGAGTGGACCGCGCACCGCGGCGCAGAGCATGTGTGATGATGCG 1370  
 Qy 8710 GACCTGTGATGAGTGGTGAACGAGCGCACCAACATGTCTGTGAGACCACTTCTGGGAT 8769  
 Db 1371 GAGCTGTGATGAGTGGTGAACGAGCGCACCAACATGTCTGTGAGACCACTTCTGGGAT 1430  
 Qy 8770 GCAGACCACTTCAACCGCGACTTCTTTCGCGCTGATGAGCGCGGACTGCTCATGACGG 8829  
 Db 1431 GCAGACCGTGTCAACCGCGACTTCTTTCGCGCTGATGAGCGCGGACTGCTCATGACGG 1490  
 Qy 8830 GATGTGAGTGTGCTGCTGCTCTGAGCTGCGGACCTGATGCTGTGAGAGCGCTGACTT 8889  
 Db 1491 CATGTCTGTGCTGCGGAGAGTGGCGCGCGGACTGTGATCTGTGGAGCGCTCACTT 1550  
 Qy 8890 CGCGCGCGCGATGCGCGCGCGCGGCTCAACCGGACCCCGGACGCGGATGTGGGGTCC 8949  
 Db 1551 CGCGCGCGCGATGCGCGCGCGCGGCTCAACCGGACCCCGGACGCGGATGTGGGGTCC 1610  
 Qy 8950 GGAAGTGTGCGCACCGCGGCGCGGAGAGCTTCTGACGACTGTGAGCCCAACGAGAGTGA 9009  
 Db 1611 GGAAGTGTGCGCACCGCGGCGCGGAGAGCTTCTGACGACTGTGAGCCCAACGAGAGTGA 1670  
 Qy 9010 GCACCGGAGAGATCC 9024  
 Db 1671 GCACCGGAGAGGCCC 1685

RESULT 10  
 US-09-105-537-7/c  
 ; Sequence 7, Application US/09105537A  
 ; Patent No. 6265202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600 438US1  
 ; CURRENT APPLICATION NUMBER: US/09/105,537A  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 7  
 ; LENGTH: 1248  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-105-537-7

Query Match 6.7%; Score 595.4; DB 3; Length 1248;  
 Best Local Similarity 69.2%; Pred. No. 6,9e-89;  
 Matches 812; Conservative 0; Mismatches 361; Indels 0; Gaps 0;  
 Qy 3926 CGGTTGATCAACGACCGCTGTGTGAGCGAGTGTGATGATGATGATGATGATGAT 3985  
 Db 1173 CGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1114  
 Qy 3986 GTTCTCACTGACACCGAGGCGCGGTGCGGAGGCGGAGACACCGTTGCGGAGCACTC 4045  
 Db 1113 GTTCTCACTGACACCGAGGCGCGGTGCGGAGGCGGAGACACCGTTGCGGAGCACTC 1054  
 Qy 4046 GGTGTGTGACCGCGGAGCGGTGCTGCGGTGAGGAGGCTGATGTGTGAGAGCGGCG 4105



Db 1053 GGTGTGCGGAGCGGGGCTGCGGGCTGCGCGGTAACGCTCCAGTCTGTGCGACCCGG 994  
Qy 4106 GAGGAAGTAGGCGCTGTGCGAGACCTTCTCCGCGCGAGAGATGCCCTGCGACCGCGTACG 4165  
Db 993 CGAAGATAGGCGCGGGTGTGCGACGCTTGGGCTTTAGAACCTTCATGACGAGGTGCGG 934  
Qy 4166 GTGGATCCGAGTGGCGGTGCGCTCACAGAGATGATCACTGATGATGCTCTCTC 4225  
Db 933 GTGGATCCGAGTGGCGGTGCGCTCTGCTGATCTGACATCACTGATGATGCTGTTGTAAGCC 874  
Qy 4226 GTGGGCGGAGCGAGTGCACGATGACCGCGGCTGATCTCGCGAGCTCTGCTGTAGAG 4285  
Db 873 GTGGCGGTGCTGTGCTGCGCGAGAGAGACGCGGGAGAGTCCGCGAGTGTCTCGCGGTAGGC 814  
Qy 4286 CGCGTGTGATCCCGGTGTGCTCTGCTGTGCTCGCGGAAAGCGCTTCGAGGAGGTGAGGCC 4345  
Db 813 GGGCTGTGTGCGCGGTTCGCGTGTGATGACCTCGGAGAAAGCGCTTCGAGGAGGTGAGGCC 754  
Qy 4346 CATGGCGCGGCGGCACTCGCTCATCTTGTGCGGTGAGTCCGATCTCGGTGACCACTTGTG 4405  
Db 753 CATGGCGGCGGCGGCTCGCTCATCTTGTGCGGTGAGTCCGCGGCGGCGGTGCGCGCGG 694  
Qy 4406 CGGCGCGATGCGGAGTGTGATGATGCGCGGATCGCTTCGCGCGAGAGCGCGTGTGCTG 4465  
Db 693 CAGGTGCGAGCGGAGTGTGAGAGGCGCGGATCGCGGCGGCGAGGTGCGCGGTGCTGCTG 634  
Qy 4466 GACCAACGCGCGCGCTCTGAGAGGCGGTGACCGCGCTTGTGTGCGGTGAGAGCTGAACTTC 4525  
Db 633 GACACGCGCGCGCGCTCTGAGAGGCGGTGACCGCGCTTGTGTGCGGTGAGAGCTGAACTTC 574  
Qy 4526 GGGCTGACCGGATCGCGCCACCGGACGTGACCGCGTGTGAGGCGAGGCGGTGCGCGG 4585  
Db 573 GGGCTGCGCGAGCTGTGCGCGGCGGCGCGCGCTGTGACCGCGAGCGAGCGGCGGTGCGCGG 514  
Qy 4586 GTGCAAGAGAGCTTGAACCTGTGTGTGCGCGGAGTCTTGTGCGCGAGCGCTTCAAGGTG 4645  
Db 513 GTGCAAGTACAGCGCGAGCGGTGTGCTGTGCGCGAGCTTGTGCGAGCTGTGCGCGGCGCA 454  
Qy 4646 TGTGTGCGCGCCACAGGTGATGCGCGAGATGCGCGCGGTGTGCGCGGCTGTGAGCGCGG 4705  
Db 453 GGGGCGCGCCGAGGTGAGCGCGGACCGCGCGAGGTGTGCGCGGTGTGAGCGCGCGGCG 394  
Qy 4706 GACGTGTGCGGGTGCACAGACCGGATGCGCGGCTGTGCGAGGACCGCGGTGTGAG 4765  
Db 393 CACGTGTGCGGGTGCAGGTGTGCGCGGTGTGCGAGTGTGCGAGGAGACCGCGGTGTGAG 334  
Qy 4766 TCCGAGCGAGCTGCGCGGTGTGCGCGGTGTGCGAGGTGTGATGAGCGGATGTATCTTC 4825  
Db 333 GCGGATTCAGAGCGAGTGTGCGGTGTGCGCGGCGGAGCGTGTGAGCGGATGTATCTTC 274  
Qy 4826 ACCGTGTGACGTACCGGCGCGGACGACGATTCAGAGGCGAGCGGTGTGCGAGGTGCG 4885  
Db 273 GCGGTGTGAGCGCGGCGGTGTGCGAGGAGCTGAGAGCGCGCGGTGTGCGAGGTGCG 214  
Qy 4886 GATGCAAGTGCCTTACCGCGAGCGAGGTGCGCGGCGGCTGTGAGTCCCGGACGAGGAG 4945  
Db 213 CAGCGCATGCGCGAGCTCGCGAGCGCGCGGCGGCGGCTGTGAGTCCCGGAGCGAGGCGG 154  
Qy 4946 CCGCGGTGTGAGCGGAGTGTGATGAGGCGGCTTCAGAGCGGCGCGGCGGAGAGCGCTG 5005  
Db 153 GCGCGGTGTGAGCGGAGCTGTGCGAGGCGCGCGGTGTGAGCGGCTGTGAGCGCTGAGC 94  
Qy 5006 CCGGTGTGCGGATGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 5065  
Db 93 GCGGTGTGAGCGGAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 34  
Qy 5066 GAAAGTCCGAGATGCGGTGTGAGGCGGCTTTCAC 5098  
Db 33 GAAAGATGCGAGGTGTGAGGCGGCTTTCAC 1

Sequence 21, Application US/09320878A  
Patent No. 6117659  
GENERAL INFORMATION:  
APPLICANT: ASHLEY, Gary  
APPLICANT: BETLACH, Melanie C.  
APPLICANT: BETLACH, Mary C.  
APPLICANT: MCDANIEL, Robert  
APPLICANT: TANG, Li  
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
FILE REFERENCE: 300622002120  
CURRENT APPLICATION NUMBER: US/09/320,878A  
EARLIER FILING DATE: 1999-05-27  
EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
EARLIER FILING DATE: 1998-08-28  
EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
EARLIER FILING DATE: 1998-05-06  
EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
EARLIER FILING DATE: 1997-04-30  
EARLIER APPLICATION NUMBER: 60/119,139  
EARLIER FILING DATE: 1999-02-08  
EARLIER APPLICATION NUMBER: 60/100,880  
EARLIER FILING DATE: 1998-09-22  
EARLIER APPLICATION NUMBER: 60/087,080  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
LENGTH: 5970  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-320-878-21  
Query Match  
Query Similarity 5.8%; Score 519.2; DB 3; Length 5970;  
Best Local Similarity 70.1%; Pred. No. 2.5e-76;  
Matches 698; Conservative 0; Mismatches 298; Indels 0; Gaps 0;  
Qy 4104 GGGAGAGATGAGGCTGTGCGACGACCTTCTCGCGCGGAGATGCGCTGACCGCGTCA 4163  
Db 1 GGGAGAGATGAGGCGGAGTGTGCAAGCGCTTCCGCTTACAGAACCTCCATGACAGAGTGC 60  
Qy 4164 CGGTGATGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4223  
Db 61 CGGTGATGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 120  
Qy 4224 TCGTGTGCGGAGAGAGTGCAGGTGACGCGGTGAGCGCGGTGAGCGCGGTGAGCGCGGT 4283  
Db 121 CGGTGTGCGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 180  
Qy 4284 AGCGGTGTGAGCGCGGTGTGCTCCGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4343  
Db 181 SCGGGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 240  
Qy 4344 CCGATGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4403  
Db 241 CCGATGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 300  
Qy 4404 TCCGCGCGGATGCGGAGTGTGATGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 4463  
Db 301 GCGAGGTGAGAGCGGAGTGTGAGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 360  
Qy 4464 GTGACACGCGCGCGCTTCAGAGCGGTGACCGGCTTGTGTGTGAGCGGTGAGAGCGGT 4523  
Db 361 GTGACAGAGGCGCGCGCTTCAGAGCGGTGAGAGCGGTGAGAGCGGTGAGAGCGGTGAG 420  
Qy 4524 TCGGTGTGACCGGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 4583  
Db 421 TCGGTGTGCGGAGCTTCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 480  
Qy 4584 GCGTGTGAGAGAGAGTGTGACCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 4643  
Db 481 GCGTGTGAGAGAGAGTGTGACCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 540  
Qy 4644 GCTGTGCGCGCCACAGGTGTACGCGGAGATGCGCGCGGTGCGCGGTGAGCGAGCGG 4703



```

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-21

Query Match      5.8%; Score 519.2; DB 3; Length 5970;
Best Local Similarity 70.1%; Pred. No. 2.5e-76;
Matches 698; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

QY 4104 GGGGGAAGTAGGAGCTGTGCGACGACCTTCTCCGCGGACGATCGCTGACGCGGTCA 4163
DB 1 GGCAGAAAGTAGGCGCGGAGTGTGACACGCTTCCGCTTCAGAACTTCATAGCAAGGTG 60
QY 4164 CGGTGATGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4223
DB 61 CGGTGATGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 120
QY 4224 TCGTTCGCGGCGGAGCGAGTGCACGTGACCGCGGTGACGTGCGGCACTCGCTGTAG 4283
DB 121 CCGTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 180
QY 4284 AGCGCGTGTGACCCGGTGTGCTCCCTGCTCTCGGCGAAACGCGTGCAGGAGTACG 4343
DB 181 SCGGCGTGTGCTCCGCTGCTCCGCTGCTCCGCTGCTCCGCTGCTCCGCTGCTCCG 240
QY 4344 CCATGCGCGGCGGCGACCTGCTCATCTTGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4403
DB 241 CCATGCGCGGCGGCGGCTGCTCATCTTGTGCGGTGCGGTGCGGTGCGGTGCGGTG 300
QY 4404 TCGGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 4463
DB 301 GGCAGGTGCGAAAGCGAAGTGTGTGAGGCGCGGAGTCCGCGGCGGCGGCGGTGCG 360
QY 4464 GTGACCAACGCGCGCGCTGCGAAGCGGTGACCGGCTTGTGTGCGGTGCGGTGCGGTG 4523
DB 361 GTGACCAACGCGCGCGCTGCGAAGCGGTGACCGGCTTGTGTGCGGTGCGGTGCGGTG 420
QY 4524 TCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4583
DB 421 TCGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 480
QY 4584 GCGTTCGAAAGAGCTTACCTGCTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4643
DB 481 GCGTTCGAAAGAGCTTACCTGCTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 540
QY 4644 GCTGTGCGCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4703
DB 541 CAGGCGCGCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 600
QY 4704 GCGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4763
DB 601 GCGAGTGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 660
QY 4764 AGTTCGAGGCGAGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4823
DB 661 AGTTCGAGGCGAGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 720
QY 4824 TCACCGGTGAGCGTGCACCGGCGCGGCGGACGACGCTTCAGGCGGCGGCGGTGAGG 4883
DB 721 TCACCGGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 4884 GCGATGCAGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4943
```

```

DB 781 GCGACGCGCAAGCGCGGAGCCCGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 4944 GCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 5003
DB 841 GCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 900
QY 5004 TCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 5063
DB 901 GCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 960
QY 5064 CCGAAGATGCGGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 5099
DB 961 CCGAAGATGCGGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 996

RESULT 14
US-09-793-708-21
; Sequence 21, Application US/09793708
; Patent No. 6902913
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MEDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: 4139 and 4819
; LOCATION: unsure
; OTHER INFORMATION: unsure of nucleotides at these positions
US-09-793-708-21

Query Match      5.8%; Score 519.2; DB 3; Length 5970;
Best Local Similarity 70.1%; Pred. No. 2.5e-76;
Matches 698; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

QY 4104 GGGGGAAGTAGGAGCTGTGCGACGACCTTCTCCGCGGACGATCGCTGACGCGGTCA 4163
DB 1 GGCAGAAAGTAGGCGCGGAGTGTGACACGCTTCCGCTTCAGAACTTCATAGCAAGGTG 60
QY 4164 CGGTGATGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4223
DB 61 CGGTGATGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 120
QY 4224 TCGTTCGCGGCGGAGCGAGTGCACGTGACCGCGGTGACGTGCGGCACTCGCTGTAG 4283
DB 121 CCGTTCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 180
QY 4284 AGCGCGTGTGACCCGGTGTGCTCCCTGCTCTCGGCGAAACGCGTGCAGGAGTACG 4343
DB 181 SCGGCGTGTGCTCCGCTGCTCCGCTGCTCCGCTGCTCCGCTGCTCCGCTGCTCCG 240
```

QY 4344 CCATGCGCGGCGCACTGCTCATCTTGCCGTTGGCGCATCTCGGTGACCACTTG 4403  
 Db 241 CCATGCGCGGCGGCGGCTGCTCATCTTGCGGTTGGCTCCGCGGCGGCGGCTGCGCG 300  
 QY 4404 TCGGCGCGCATGCGGAACTTGTGCATGAGCCCGATCCGTTCCGCGCAGGCGCTGCTG 4463  
 Db 301 GCGAGGTGCAAGCGGAAGTTGTGAGGCGCGGATCCGCGGCGGCGGAGGTGCGGCTGCTG 360  
 QY 4464 GTACCAACCCCGCCCGCTTGAAGGCGGTGACCGCTTGGTGGGTGGAAGCTGAACAC 4523  
 Db 361 GTACCAACCGCGCGCTTGAAGGCGGTGACCGCTTGGTGGGTGGAAGCTGAACAC 420  
 QY 4524 TCGGCGCGCATGCGGATCCGCGCAACCGGATCCGCTGCGGCGCAGGCGGTGCGG 4583  
 Db 421 TCGGCGCGCGCGAGGCTGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGG 480  
 QY 4584 GCGTCGAAGAAGCTTGAAGCTTGTGCTGCGGCGGCGGATCTTTCGCGCAGCGCTTCAAGT 4643  
 Db 481 GCGTCGAAGAAGCTTGAAGCTTGTGCTGCGGCGGCGGATCTTTCGCGCAGCGCTTCAAGT 540  
 QY 4644 GCTGCTGCGGCGCGCAAGGTGAGCGGCGGATGCGGCGGCTGCGGCGGCGGCGGCGG 4703  
 Db 541 GCGGCGCGGCGCGCGAGGTGAGCGGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCG 600  
 QY 4704 GCGACGTGCTCGGCGGTGCAACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4763  
 Db 601 GCGACGTGCTCGGCGGTGCAACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 QY 4764 AGTCGCAAGCGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4823  
 Db 661 AGTCGCAAGCGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
 QY 4824 TCAACGCTGACGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4883  
 Db 721 TCAACGCTGACGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
 QY 4884 GCGATGCAAGCGGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4943  
 Db 781 GCGATGCAAGCGGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 QY 4944 GCGCGCGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5003  
 Db 841 GCGCGCGCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 QY 5004 TCGCGGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5063  
 Db 901 TCGCGGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 QY 5064 CCGAAGATGCGCGGATCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5099  
 Db 961 CCGAAGATGCGCGGATCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 996

## RESULT 15

US-09-758-759-30/c  
 ; Sequence 30, Application US/09758759  
 ; Patent No. 6861513  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hosted, Thomas J.  
 ; APPLICANT: Wang, Tim X.  
 ; APPLICANT: Horan, Ann C.  
 ; TITLE OF INVENTION: Evernimicin Biosynthetic Genes  
 ; FILE REFERENCE: ID0983K US  
 ; CURRENT APPLICATION NUMBER: US/09758,759  
 ; CURRENT FILING DATE: 2001-01-11  
 ; PRIOR APPLICATION NUMBER: US 60/175,751  
 ; NUMBER OF SEQ ID NOS: 204  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 30  
 ; LENGTH: 1452  
 ; TYPE: DNA

ORGANISM: Micromonospora carbonacea

; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1452)  
 ; OTHER INFORMATION: evic  
 US-09-758-759-30

Query Match 5.2%; Score 461.2; DB 3; Length 1452;  
 Best Local Similarity 60.8%; Pred. No. 7.1e-67;  
 Matches 820; Conservative 0; Mismatches 498; Indels 30; Gaps 3;

QY 5115 AGACGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5174  
 Db 1450 AGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1391  
 QY 5175 CGACGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 5234  
 Db 1390 GGAAGAGGCGCTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1331  
 QY 5235 CGAAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5294  
 Db 1330 GGAAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1271  
 QY 5295 CCTCTCGGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5354  
 Db 1270 CCTCTCGGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1211  
 QY 5355 CGAAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5414  
 Db 1210 CGAAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1163  
 QY 5415 CGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5474  
 Db 1162 CGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1103  
 QY 5475 CGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5534  
 Db 1102 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1043  
 QY 5535 AGACGCGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 5594  
 Db 1042 AGACGCGGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 983  
 QY 5595 AGATGCTGAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5654  
 Db 982 CCAAGGAGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 923  
 QY 5655 TCGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5714  
 Db 922 GCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 869  
 QY 5715 CCGAAGATGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 5774  
 Db 868 CCGAAGATGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 809  
 QY 5775 AAGCAGCAGTGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5834  
 Db 808 GCAAGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 761  
 QY 5835 CCGCGGCTTCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5894  
 Db 760 CCGCGGCTTCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701  
 QY 5895 CGTGTGCTGAGCGGAGCATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5954  
 Db 700 GCGCGCTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 641  
 QY 5955 GCTCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6014  
 Db 640 CCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 581  
 QY 6015 AGGAACCTGCTCGGACTGGAAGAGTGGAGGATCCGCTGCGGCGGCGGCGGCGGCGGCGGCGG 6074

```
Db      580 AGGAGCCCTGCTCCGACTGCAAGGACGTCAACCAAGCACCCGACCCGCCCGGCGGCTGA 521
QY      6075 AGTACTCGATGAACTGGAACGGGGATCCGTTGGAACCTCATGTAGTTGCTGGGGTGG 6134
Db      520 AGTAGTCGAGGTACTTGTGTCGCCGCCGACACCGGCTGTAGTTGCTCGGGTGG 461
QY      6135 CCTGCAGGGTTCGGCGAAGCTGGAAGGCGTTGATGTTGCCAGGTTCCGGCTTGGCTTGA 6194
Db      460 CTTGCACCGTTCGGGGAAGCTGCAAGAGTTGATGTTGCCGGGCTCCACCTTGGCTTGA 401
QY      6195 CCAAGCGTGCAGTACGCCCTCGAATCCCGCACGATCAGACGAGGAACCCGATCTCGG 6254
Db      400 GCAAGCAGTACGGTGTCCCGTCGACGACTTGAACGACATGCGGAGATGCCGATCTCCG 341
QY      6255 GTTGAAGATGATGGTTGATCCAGTCCCGTCCGATCCGAAGTTGTCGGACGTGA 6314
Db      340 GCTGTTGATGATGGCTGTGTGCTGATTCGCGCACCGCGCGGTAGTGTGTGACGTGA 281
QY      6315 GGCCTCGATGAGAGAAAGCGCCGGAGTGTGCGCCAGCCGATCCTCCGGGTGA 6374
Db      280 GCGCTCGATCAGAAAGAACCGGCGGCTGTGTGCCGAGGTTGCCGGTCAACGGGTGA 221
QY      6375 ACGACCAAGCTTCCATGTGCTGAAGGCACTCGGTGCACCTCGAGCCGATGCTCGGCG 6434
Db      220 ACGCCCAACCGGCGAGCGGTCAAGCGGCAAGCGGTCCACCGGCAAGTGTGACCGGG 161
QY      6435 CGGGTGGGCCAACGATCGTGAAGTC 6462
Db      160 TCCGCTCGGCGAACCAAGAGGAAGTC 133
```

Search completed: April 7, 2006, 02:59:03  
Job time : 1490.43 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 7, 2006, 01:33:46 ; Search time 5144.9 Seconds  
(without alignments)  
11689.676 Million cell updates/sec

Title: US-10-611-442-1

Perfect score: 8942

Sequence: 1 gcgcgcctgcacacacatgcg.....gtgagacacggagagatcc 9024

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapekt 1.0

Searched: 4996997 seqs, 333346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_21:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001as:\*
- 5: geneseqn2001bs:\*
- 6: geneseqn2002as:\*
- 7: geneseqn2002bs:\*
- 8: geneseqn2003as:\*
- 9: geneseqn2003bs:\*
- 10: geneseqn2003cs:\*
- 11: geneseqn2003ds:\*
- 12: geneseqn2004as:\*
- 13: geneseqn2004bs:\*
- 14: geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8942	100.0	9024	12	AD114147
2	3037.4	34.0	8160	2	AAZ25774
3	3032.6	33.9	8051	2	AAZ72685
4	2606.8	29.2	47961	4	AAZ30757
5	1309.4	14.6	12441	3	AAZ87284
6	1309.4	14.6	13613	6	AAZ39043
7	1309.4	14.6	13613	12	ADL91931
8	1307.8	14.6	13613	3	AAZ87319
9	1241.2	13.9	60196	10	AAZ55810
10	1203.2	13.5	84428	12	ADW45913
11	729.4	8.2	1565	3	AAZ75638
12	729.4	8.2	1565	3	AAZ56006
13	729.4	8.2	1565	8	ADA09423
14	729.4	8.2	1565	10	ADH53467
15	729.4	8.2	1565	10	ABZ56095
16	722	8.1	1458	6	AAZ87287
17	722	8.1	1458	6	AAZ39045
18	722	8.1	1458	12	ADL91937
19	720.6	8.1	85915	13	ADL15447

C	20	711.6	8.0	3412	2	AAZ25772	AAZ25772 S. erythra
C	21	701.4	7.8	3756	2	AAZ72684	AAZ72684 Sugar bio
C	22	667.2	7.5	1467	10	AAZ55828	AAZ55828 Micromono
C	23	612.8	6.9	17596	12	AD114148	AD114148 M. megal
C	24	596.6	6.7	1227	10	AAZ55827	AAZ55827 Micromono
C	25	595.4	6.7	1248	3	AAZ87286	AAZ87286 S. venez
C	26	595.4	6.7	1248	6	AAZ39044	AAZ39044 Streptomy
C	27	595.4	6.7	1248	12	ADL91935	ADL91935 Streptomy
C	28	519.2	5.8	5970	3	AAZ75635	AAZ75635 Nucleotid
C	29	519.2	5.8	5970	3	AAZ56003	AAZ56003 Contig 00
C	30	519.2	5.8	5970	8	ADA09420	ADA09420 S. venez
C	31	519.2	5.8	5970	10	ADH53464	ADH53464 S. venez
C	32	519.2	5.8	5970	10	ABZ56092	ABZ56092 Contig 00
C	33	500.4	5.6	45055	10	ABZ66808	ABZ66808 Orthosomy
C	34	496.6	5.6	1401	10	ABZ66869	ABZ66869 Orthosomy
C	35	474.2	5.3	1404	8	ABZ37554	ABZ37554 Streptomy
C	36	474.2	5.3	59816	8	ABZ37515	ABZ37515 Streptomy
C	37	471.2	5.3	59816	8	ABZ37515	ABZ37515 Streptomy
C	38	463.4	5.2	30943	12	ADH80034	ADH80034 Spiramyl
C	39	463.4	5.2	30943	12	ADN97550	ADN97550 S. ambofac
C	40	463.4	5.2	30943	12	ADN97550	ADN97550 S. ambofac
C	41	461.2	5.2	1455	10	ABZ66691	ABZ66691 Orthosomy
C	42	461.2	5.2	37116	10	ABZ66810	ABZ66810 Orthosomy
C	43	461.2	5.2	109519	5	AAZ508693	AAZ508693 Micromono
C	44	451.2	5.0	1410	12	ADN80073	ADN80073 Spiramyl
C	45	451.2	5.0	1410	12	ADN97589	ADN97589 S. ambofac

## ALIGNMENTS

RESULT 1	AD114147	AD114147 standard; DNA; 9024 BP.
ID	AD114147	AD114147
XX	AD114147	AD114147
AC	AD114147	AD114147
XX	AD114147	AD114147
DT	15-APR-2004	(first entry)
XX	15-APR-2004	(first entry)
DE	M. megalomicea coemid PKOS079-138B SEQ ID NO:1.	
XX	M. megalomicea coemid PKOS079-138B SEQ ID NO:1.	
KM	db: polyketide; enzyme; MegR; MegF; MegG; MegCIV; MegCV; MegBVI; MegBIII;	
KM	MegL; MegM; megosamine; megalomycin.	
XX	Micromonospora megalomicea.	
OS	Micromonospora megalomicea.	
XX	WO2004003169-A2.	
PN	WO2004003169-A2.	
PD	08-JAN-2004.	
XX	08-JAN-2004.	
XX	30-JUN-2003; 2003WO-US020681.	
XX	30-JUN-2003; 2003WO-US020681.	
XX	28-JUN-2002; 2002US-0393016P.	
XX	28-JUN-2002; 2002US-0393016P.	
PA	(KOSA-) KOSAN BIOSCIENCES INC.	
XX	(KOSA-) KOSAN BIOSCIENCES INC.	
XX	Hutchinson RC, Katz L, Reid R, Hu Z, Gramajo H;	
XX	Hutchinson RC, Katz L, Reid R, Hu Z, Gramajo H;	
DR	WPI; 2004-203379/19.	
XX	WPI; 2004-203379/19.	
PT	Novel isolated, purified, or recombinant nucleic acid comprising	
PT	polyketide modifying gene, there gene encodes polyketide modifying enzyme	
PT	e.g., MegR, MegK, or MegM enzymes useful for producing modified	
PT	polyketide.	
XX	Example 2; SEQ ID NO 1; 51pp; English.	
XX	Example 2; SEQ ID NO 1; 51pp; English.	
CC	The invention relates to a novel isolated, purified, or recombinant	
CC	nucleic acid (1) comprising a polyketide modifying gene, where the gene	
CC	encodes a polyketide modifying enzyme chosen from MegR, MegF, MegG,	
CC	MegCIV, MegCV, MegBVI, MegBIII, MegL, and MegM enzymes. A method of the	
CC	invention is useful for producing a modified polyketide, which involves	
CC	culturing a recombinant cell comprising the recombinant nucleic acid	



under conditions in which the cell expresses a product of a gene encoded by the nucleic acid under conditions in which the unmodified polypeptide is present, and producing the modified polypeptide. The cell produces megasamine and can attach megasamine to a polypeptide, where the cell, it is naturally occurring non-recombinant state cannot produce megasamine. The present sequence contains upstream megalomycin modification enzyme genes.

Sequence 9024 BP; 1324 A; 3022 C; 3127 G; 1469 T; 0 U; 82 Other;

Query Match 100.0%; Score 8942; DB 12; Length 9024;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 9024; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GCGGCTTCGATCACCATGATCGCTTAATGTCGGGTCCATGCTTCGATGGGGAT 60
DB 1 GCGGCTTCGATCACCATGATCGCTTAATGTCGGGTCCATGCTTCGATGGGGAT 60
QY 61 GTAGTCAATTGCAAAATCCGAGACCGTGTAAAGCTTCGAGTCCGCTGTTTC 120
DB 61 GTAGTCAATTGCAAAATCCGAGACCGTGTAAAGCTTCGAGTCCGCTGTTTC 120
QY 121 GTCAATTCCTCCATCGCCGCAAAAGACCGCGACCGGAAACCGAGAAATGTTCTCGCA 180
DB 121 GTCAATTCCTCCATCGCCGCAAAAGACCGCGACCGGAAACCGAGAAATGTTCTCGCA 180
QY 181 ATGCTCCCTCGTCCACCGCGACCGAGTGGTTCGGGTCTCGGTCTCGAGTCTGG 240
DB 181 ATGCTCCCTCGTCCACCGCGACCGAGTGGTTCGGGTCTCGGTCTCGAGTCTGG 240
QY 241 GACGACGAGCCGCGGTGACGAGAGGCTCACACCCCTCGACAGTACATTCGATCTGCGC 300
DB 241 GACGACGAGCCGCGGTGACGAGAGGCTCACACCCCTCGACAGTACATTCGATCTGCGC 300
QY 301 AAGATGTTCTGTGGCGGTGACCGGTCTGCGCCGCGAGAGTCACTCGAGTCTGCTATC 360
DB 301 AAGATGTTCTGTGGCGGTGACCGGTCTGCGCCGCGAGAGTCACTCGAGTCTGCTATC 360
QY 361 ACCCGGGCGCGGCTATCTGCTCGCGGTGACCGGATCGGCTCGAGTCCGAGAGTAC 420
DB 361 ACCCGGGCGCGGCTATCTGCTCGCGGTGACCGGATCGGCTCGAGTCCGAGAGTAC 420
QY 421 CAGCGTCTGATGTCCGCGCGGTGACCGCGCTCGGCTCGGCGACGAGTGAACAGGCAAC 480
DB 421 CAGCGTCTGATGTCCGCGCGGTGACCGCGCTCGGCTCGGCGAGAGTGAACAGGCAAC 480
QY 481 CGCAGACTGACCGAGCGCTCGGCTCTGTGGCGCGGGCCGCGCTGTGACGCTCCGCTG 540
DB 481 CGCAGACTGACCGAGCGCTCGGCTCTGTGGCGCGGGCCGCGCTGTGACGCTCCGCTG 540
QY 541 GGTGGGTGTGAGATCGAAGGCTGCGGACCTGAGAGTCTGTGCTCATGCGCGAGAA 600
DB 541 GGTGGGTGTGAGATCGAAGGCTGCGGACCTGAGAGTCTGTGCTCATGCGCGAGAA 600
QY 601 TACCTGTGCGCGCGAAGTGTGCTGAGGAGATGTACCGGAGAGCCCTCATCGAGCTGAC 660
DB 601 TACCTGTGCGCGCGAAGTGTGCTGAGGAGATGTACCGGAGAGCCCTCATCGAGCTGAC 660
QY 661 GCGCTCACCGCGAGAAATCCGCTGACAGAGGGCTTCAGAGCGCAGATCATGCGGCGCTG 720
DB 661 GCGCTCACCGCGAGAAATCCGCTGACAGAGGGCTTCAGAGCGCAGATCATGCGGCGCTG 720
QY 721 CATCTCAATGTGTCGACGCGCGGAGGCTTTCGACCGGTTGTGCGCAACCTC 780
DB 721 CATCTCAATGTGTCGACGCGCGGAGGCTTTCGACCGGTTGTGCGCAACCTC 780
QY 781 GTGACGAACTGGGTCTGGAACCGAGCGCAGGTGCAACGATCCACAGGCGATCTCTG 840
DB 781 GTGACGAACTGGGTCTGGAACCGAGCGCAGGTGCAACGATCCACAGGCGATCTCTG 840
QY 841 AACGCGGAGACCGACTTCGAGAGACGATCTGCGCGTCACTCGTCCGTTCCGTCGAGGTC 900
DB 841 AACGCGGAGACCGACTTCGAGAGACGATCTGCGCGTCACTCGTCCGTTCCGTCGAGGTC 900
```

```
QY 901 GCCGCAAGATTGGGTGGGTCCGATCCGGGTCCGGCGAGCTGACCGATTACCGCTACGCG 960
DB 901 GCCGCAAGATTGGGTGGGTCCGATCCGGGTCCGGCGAGCTGACCGATTACCGCTACGCG 960
QY 961 GACACCTTGAGCCGCAATCAACGACATTTGGCCAAAATTCGACATCTGTGCCCCGGGGGGA 1020
DB 961 GACACCTTGAGCCGCAATCAACGACATTTGGCCAAAATTCGACATCTGTGCCCCGGGGGGA 1020
QY 1021 CCGGTTGAGAGAGAACGATGAGGAGAACATGACACTATATGAAAGATCCGAGCATG 1080
DB 1021 CCGGTTGAGAGAGAACGATGAGGAGAACATGACACTATATGAAAGATCCGAGCATG 1080
QY 1081 GCCGAGAGGCGGTTCTGCTGACCTGCTGGCGGTGATGCGCGACAGGCAACCGGTCTGG 1140
DB 1081 GCCGAGAGGCGGTTCTGCTGACCTGCTGGCGGTGATGCGCGACAGGCAACCGGTCTGG 1140
QY 1141 CAGGACCAATGACGGGTCTGGACATTTCCGCCACAGTGAAGTACCCGAGAGTCTCCG 1200
DB 1141 CAGGACCAATGACGGGTCTGGACATTTCCGCCACAGTGAAGTACCCGAGAGTCTCTCCG 1200
QY 1201 GACACCGCACTTCTCTCGACCCGACCCGCTGATCGAGGGGGCGACCCGACGCG 1260
DB 1201 GACACCGCACTTCTCTCGACCCGACCCGCTGATCGAGGGGGCGACCCGACGCG 1260
QY 1261 GGGATGATCCACAGATCGACCCGCGAGAGCACCGGCGCCTTGGCAAAGTGTGACAGC 1320
DB 1261 GGGATGATCCACAGATCGACCCGCGAGAGCACCGGCGCCTTGGCAAAGTGTGACAGC 1320
QY 1321 GCTTCAACCCCGGTACGATGCGGACCTCGAACCGGGACATCCGGAGGTGACCCCGTGG 1380
DB 1321 GCTTCAACCCCGGTACGATGCGGACCTCGAACCGGGACATCCGGAGGTGACCCCGTGG 1380
QY 1381 CTGCTGGCGGACGCGCGGTGACCGGCTGACCTGTGAGAGGCGCTCGCTTCCGCTGCG 1440
DB 1381 CTGCTGGCGGACGCGCGGTGACCGGCTGACCTGTGAGAGGCGCTCGCTTCCGCTGCG 1440
QY 1441 GTCAAGATCGTGGCGGAGCTGTGGGTGGGTGCGGATGGAACAACAGATTGCTGAC 1500
DB 1441 GTCAAGATCGTGGCGGAGCTGTGGGTGGGTGCGGATGGAACAACAGATTGCTGAC 1500
QY 1501 TGGTCCGCGCGCTGTGTGACATCAATGAGACGACCGGATCCGACCTGTGCA 1560
DB 1501 TGGTCCGCGCGCTGTGTGACATCAATGAGACGACCGGATCCGACCTGTGCA 1560
QY 1561 CGCATCATGACGAGGTGGAACCGGCTACCTGACCTGCTGACAGGTGTCCGGAACG 1620
DB 1561 CGCATCATGACGAGGTGGAACCGGCTACCTGACCTGCTGACAGGTGTCCGGAACG 1620
QY 1621 CGGCGGACCCCGGAGACGACTGATCTCCGGGTGGGTGCTGGCGAGGTGACGCGGCG 1680
DB 1621 CGGCGGACCCCGGAGACGACTGATCTCCGGGTGGGTGCTGGCGAGGTGACGCGGCG 1680
QY 1681 ACCCTCGACGAGTGAAGCGGCGCACTTCTGACAGGTGCTGTGCGGGGCACTC 1740
DB 1681 ACCCTCGACGAGTGAAGCGGCGCACTTCTGACAGGTGCTGTGCGGGGCACTC 1740
QY 1741 ACCACCAACGCTGCTGGGGAACATGCTCGGACCCGCGACCGGAGATGAGCTGG 1800
DB 1741 ACCACCAACGCTGCTGGGGAACATGCTCGGACCCGCGACCGGAGATGAGCTGG 1800
QY 1801 ACCGCGCGCGAGAGACCCGGGTCTGATCGCGCGATCATGAGAGGTGTGGTTTC 1860
DB 1801 ACCGCGCGCGAGAGACCCGGGTCTGATCGCGCGATCATGAGAGGTGTGGTTTC 1860
QY 1861 CGCCCGCGGTTCCCGATGACGACCAACGACGAGGCGACCACTGTGGTGGGTTC 1920
DB 1861 CGCCCGCGGTTCCCGATGACGACCAACGACGAGGCGACCACTGTGGTGGGTTC 1920
QY 1921 GAGATCCCGGCGAGTGAATGATGAAACATGGGGGTCTGGGCGCAACCGGATCCGCTG 1980
DB 1921 GAGATCCCGGCGAGTGAATGATGAAACATGGGGGTCTGGGCGCAACCGGATCCGCTG 1980
```



QY 1981 GCGCATCCGACCCGAGACAGCTTGCACCCGTCGCCGAGATCGGTGTGTCGCGCAGCTC 2040  
Db 1981 GCGCATCCGACCCGAGACAGCTTGCACCCGTCGCCGAGATCGGTGTGTCGCGCAGCTC 2040  
QY 2041 TCCCTTCGGGCGACGGCGGTGCACTTCTGTCTCGGTGCCCCGCTGGGCGCGCTTGAAGAACAG 2100  
Db 2041 TCCCTTCGGGCGACGGCGGTGCACTTCTGTCTCGGTGCCCCGCTGGGCGCGCTTGAAGAACAG 2100  
QY 2101 GTCGCCCTGAGAGATCATCGCCGGGTACGGTGCACCTGGCGGTGCACCGGACGAC 2160  
Db 2101 GTCGCCCTGAGAGATCATCGCCGGGTACGGTGCACCTGGCGGTGCACCGGACGAC 2160  
QY 2161 ACGCTGCGTCACTTGCACAGATCGTCTCGGCAACCGGACGACCTCCGGGTCTGGCGCG 2220  
Db 2161 ACGCTGCGTCACTTGCACAGATCGTCTCGGCAACCGGACGACCTCCGGGTCTGGCGCG 2220  
QY 2221 GTCACCCCGCGCGAGTCCGCTGAACCCCTTGCCTCGACGCGCGGNNNNNNNNNNNN 2280  
Db 2221 GTCACCCCGCGCGAGTCCGCTGAACCCCTTGCCTCGACGCGCGGNNNNNNNNNNNN 2280  
QY 2281 NNN 2340  
Db 2281 NNN 2340  
QY 2341 NNN 2400  
Db 2341 NNN 2400  
QY 2401 CAGGCCCGGATGCTGATCCCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2460  
Db 2401 CAGGCCCGGATGCTGATCCCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2460  
QY 2461 GGTCACGGAAGTGTGCAAGCCGTGAGAGAGAACTCCCTCCCGCGGTGCGGATGAGTGTCT 2520  
Db 2461 GGTCACGGAAGTGTGCAAGCCGTGAGAGAGAACTCCCTCCCGCGGTGCGGATGAGTGTCT 2520  
QY 2521 GCGGCCCGGATGCTGATCCCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2580  
Db 2521 GCGGCCCGGATGCTGATCCCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2580  
QY 2581 GCGGCCCGGATGCTGATCCCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2640  
Db 2581 GCGGCCCGGATGCTGATCCCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2640  
QY 2641 GAGCTCTCGAGAGAGTGTGCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2700  
Db 2641 GAGCTCTCGAGAGAGTGTGCAACCGGTGCGGCGACGTCCGTTGAGTTGTTGAGCGGCG 2700  
QY 2701 TTTCCGCGCGGATCCGCAACAGCTGGGCGTGCACCCCGGACGAGGCTCTCCAAAGCGTA 2760  
Db 2701 TTTCCGCGCGGATCCGCAACAGCTGGGCGTGCACCCCGGACGAGGCTCTCCAAAGCGTA 2760  
QY 2761 ACCGAGGTGCGTGTGCAATGCGCGGAGTCCGCTCGGCGGCGTGTGAGAACCGGCGAG 2820  
Db 2761 ACCGAGGTGCGTGTGCAATGCGCGGAGTCCGCTCGGCGGCGTGTGAGAACCGGCGAG 2820  
QY 2821 GCGCTCCCGGAACTTCGGTCCGCTTCCGCTGCGACAACTTCGCTGCTCCGACGCGCTGA 2880  
Db 2821 GCGCTCCCGGAACTTCGGTCCGCTTCCGCTGCGACAACTTCGCTGCTCCGACGCGCTGA 2880  
QY 2881 GTCCTCGCGGACGGTGAAGAGTGCAGCGGCGGGTGGGCTGGGACTGTTGAGTTGAG 2940  
Db 2881 GTCCTCGCGGACGGTGAAGAGTGCAGCGGCGGGTGGGCTGGGACTGTTGAGTTGAG 2940  
QY 2941 GATGAAGTGCACAGAGTGCATGAGCGGTTGCGCGCGCGCGCGGACGAGTGTGTTGAG 3000  
Db 2941 GATGAAGTGCACAGAGTGCATGAGCGGTTGCGCGCGCGCGCGGACGAGTGTGTTGAG 3000  
QY 3001 GCGGACCGGACGCGGTGCGCGCTTCCGCGCGGACGAGTGTGTTGAGTTGAGTTGAG 3060  
Db 3001 GCGGACCGGACGCGGTGCGCGCTTCCGCGCGGACGAGTGTGTTGAGTTGAGTTGAG 3060  
QY 3061 CCGTACCGGTTGAGAGCGCGCAGCTTGCAGGTGTGAGCTGTGAGTTGAGTTGAG 3120

Db 3061 CCGTACCGGTTGAGAGCGCGCAGCTTGCAGGTGTGAGCTGTGAGTTGAGTTGAG 3120  
QY 3121 CCGGTACAGGAGGTGACAGCCGCGCGAGGTCCCAAGCGCGGCTGGCGCGACGCT 3180  
Db 3121 CCGGTACAGGAGGTGACAGCCGCGCGAGGTCCCAAGCGCGGCTGGCGCGACGCT 3180  
QY 3181 CCGGTACAGGAGGTGACAGCCGCGCGAGGTCCCAAGCGCGGCTGGCGCGACGCT 3240  
Db 3181 CCGGTACAGGAGGTGACAGCCGCGCGAGGTCCCAAGCGCGGCTGGCGCGACGCT 3240  
QY 3241 GTCGACACCAACCGCCCGATACCGCGGTTGTCAACGTTTCCAGCGCGCGGAGAGTA 3300  
Db 3241 GTCGACACCAACCGCCCGATACCGCGGTTGTGTCAACGTTTCCAGCGCGCGGAGAGTA 3300  
QY 3301 CATGCGGTGCGGTTTCTCCGACCGGACCTGTGTGATGACCGAGGTGAGATCTGTGCGC 3360  
Db 3301 CATGCGGTGCGGTTTCTCCGACCGGACCTGTGTGATGACCGAGGTGAGATCTGTGCGC 3360  
QY 3361 GGGGACCAACGAGCTCCGCGGTGCTTGTGACCTGTGACCTGTGACCGGAGAGTGAAGCG 3420  
Db 3361 GGGGACCAACGAGCTCCGCGGTGCTTGTGACCTGTGACCTGTGACCGGAGAGTGAAGCG 3420  
QY 3421 GAAATGACAGTGTGCGGCGGAGTGAAGCCGAGCTGTGACGAGGAGGCGGCTCCGCGC 3480  
Db 3421 GAAATGACAGTGTGCGGCGGAGTGAAGCCGAGCTGTGACGAGGAGGCGGCTCCGCGC 3480  
QY 3481 TACCGAGCTTCCAAACCGCGCGGCGGTTCAAGCGGCTGATGATGTTGCTCAAGTCTT 3540  
Db 3481 TACCGAGCTTCCAAACCGCGCGGCGGTTCAAGCGGCTGATGATGTTGCTCAAGTCTT 3540  
QY 3541 GCGCGCGGCGGCTTCTCCACCGCGGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 3600  
Db 3541 GCGCGCGGCGGCTTCTCCACCGCGGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 3600  
QY 3601 CAGCTGCGGACCGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 3660  
Db 3601 CAGCTGCGGACCGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 3660  
QY 3661 GAAAGCGGAGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 3720  
Db 3661 GAAAGCGGAGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 3720  
QY 3721 GGTGTGAGAGTGTGCAACCGGTGAGGCGGAGTGTGAGGCGGCTGAGGCGGCTGAG 3780  
Db 3721 GGTGTGAGAGTGTGCAACCGGTGAGGCGGAGTGTGAGGCGGCTGAGGCGGCTGAG 3780  
QY 3781 CCGCGGAGTGTGCAACCGGTGAGGCGGAGTGTGAGGCGGCTGAGGCGGCTGAG 3840  
Db 3781 CCGCGGAGTGTGCAACCGGTGAGGCGGAGTGTGAGGCGGCTGAGGCGGCTGAG 3840  
QY 3841 AGGAGTGTGCTTATGAGCGTGTGCAACCGGTGAGGCGGAGTGTGAGGCGGCTGAG 3900  
Db 3841 AGGAGTGTGCTTATGAGCGTGTGCAACCGGTGAGGCGGAGTGTGAGGCGGCTGAG 3900  
QY 3901 CCGTGTGCTTCCGCTTGTGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3960  
Db 3901 CCGTGTGCTTCCGCTTGTGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3960  
QY 3961 GATGAAGTGCACAGAGTGCATGAGCGGTTGCGCGCGCGCGCGGACGAGTGTGTTGAG 4020  
Db 3961 GATGAAGTGCACAGAGTGCATGAGCGGTTGCGCGCGCGCGCGGACGAGTGTGTTGAG 4020  
QY 4021 GAGGACCGGTTGCGGAGGCACTCGGTGTGTGTGACCGGCGGAGGCTGCTGCGGCT 4080  
Db 4021 GAGGACCGGTTGCGGAGGCACTCGGTGTGTGTGACCGGCGGAGGCTGCTGCGGCT 4080  
QY 4081 AGGAGTGTGCTGT 4140  
Db 4081 AGGAGTGTGCTGT 4140  
QY 4141 GCAGAGTGTGCTGT 4200



QY 6361 GTCTCCGGGTGAACGACGAGCTTCATGTCGTAAGGCACTCGTCACCTCGAG 6420  
Db 6361 GTCTCCGGGTGAACGACGAGCTTCATGTCGTAAGGCACTCGTCACCTCGAG 6420  
QY 6421 CCGATGCTCGGCGGCGGCGGTGAGGCAACCAATCGTGAAGTGTGCGGTGAGGAG 6480  
Db 6421 CCGATGCTCGGCGGCGGCGGTGAGGCAACCAATCGTGAAGTGTGCGGTGAGGAG 6480  
QY 6481 TCCGCGGTCGAGTGAAGAAAGTATGCGGATTTGTTGGAATTCGGAGTCGATGACC 6540  
Db 6481 TCCGCGGTCGAGTGAAGAAAGTATGCGGATTTGTTGGAATTCGGAGTCGATGACC 6540  
QY 6541 GTTACCCGATCCCGCAATACGCTCTCTCCGATGATGATGAGGCGGTCGTCGATACCG 6600  
Db 6541 GTTACCCGATCCCGCAATACGCTCTCTCCGATGATGATGAGGCGGTCGTCGATACCG 6600  
QY 6601 CCGGACTGACATTCGTGATCAGACCCCGCCAGTGTAGGGCTCCGCGCCGACGAGAG 6660  
Db 6601 CCGGACTGACATTCGTGATCAGACCCCGCCAGTGTAGGGCTCCGCGCCGACGAGAG 6660  
QY 6661 AAGGTCCGTGAACAATTCCGGGTGACCGGTCCCGCGGTGAGAAACGGGCGTGGGA 6720  
Db 6661 AAGGTCCGTGAACAATTCCGGGTGACCGGTCCCGCGGTGAGAAACGGGCGTGGGA 6720  
QY 6721 GCAACCCGATCATTCGTGATGAGAACTTCTTAATCTGTGAGGCGGACATCTTTTGAACG 6780  
Db 6721 GCAACCCGATCATTCGTGATGAGAACTTCTTAATCTGTGAGGCGGACATCTTTTGAACG 6780  
QY 6781 GTGTGTTCCGTGATGACGCGTTCGCGGCGGCTGTGAACTGTGCGTGGACCTGACCG 6840  
Db 6781 GTGTGTTCCGTGATGACGCGTTCGCGGCGGCTGTGAACTGTGCGTGGACCTGACCG 6840  
QY 6841 TTGGGCGGTGTTTTCGCGCGGTCGGAATCGCGGATTCGTGAAACGGGCGGAGGAGCG 6900  
Db 6841 TTGGGCGGTGTTTTCGCGCGGTCGGAATCGCGGATTCGTGAAACGGGCGGAGGAGCG 6900  
QY 6901 GGTGCTGACCGGATGATCTGCAATCATGAGCGCTCATGACATCTTTGTAGCATGCT 6960  
Db 6901 GGTGCTGACCGGATGATCTGCAATCATGAGCGCTCATGACATCTTTGTAGCATGCT 6960  
QY 6961 CCGGCGGAGGTCGACAGGCGCGGAAACGCGCGCATTCAGCTGTTTGAAGACGTCGA 7020  
Db 6961 CCGGCGGAGGTCGACAGGCGCGGAAACGCGCGCATTCAGCTGTTTGAAGACGTCGA 7020  
QY 7021 CATCACGTCGACAGCGCGGATGACACCGACACGACGATGTCGTCGTCGTCGTCGTCG 7080  
Db 7021 CATCACGTCGACAGCGCGGATGACACCGACACGACGATGTCGTCGTCGTCGTCGTCG 7080  
QY 7081 GTGCGCGATCAGGGAATGAGCGGTCTCATGACAGCAAGCAAGCTGTCCGGGCT 7140  
Db 7081 GTGCGCGATCAGGGAATGAGCGGTCTCATGACAGCAAGCAAGCTGTCCGGGCT 7140  
QY 7141 GCGGAAACAGCATGATCTGCTGACGCGTTCATTTGCGCGCGGACCGCTTGAAT 7200  
Db 7141 GCGGAAACAGCATGATCTGCTGACGCGTTCATTTGCGCGCGGACCGCTTGAAT 7200  
QY 7201 CCGTGCACCGGTGTCGTCGAGTGAAGTGCAGATCGCGGACCGGAGTTCGAGACGAGTGT 7260  
Db 7201 CCGTGCACCGGTGTCGTCGAGTGAAGTGCAGATCGCGGACCGGAGTTCGAGACGAGTGT 7260  
QY 7261 AAGCGATGACGAGGCTTCTGTCGTCGCGCGGACCGGTCGCACTAGGTTGAATTCGTCACG 7320  
Db 7261 AAGCGATGACGAGGCTTCTGTCGTCGCGCGGACCGGTCGCACTAGGTTGAATTCGTCACG 7320  
QY 7321 TCTTCAATCCGGAAGCTTCTATGACAGTGTGCGATGAGAGCGGCTGGGCGGGAATTTG 7380  
Db 7321 TCTTCAATCCGGAAGCTTCTATGACAGTGTGCGATGAGAGCGGCTGGGCGGGAATTTG 7380  
QY 7381 CAGATGCGCGGCTCTCTATGAGGCTTGTGCGGTCGACAGCGGATCTGTACTGATGCTC 7440  
Db 7381 CAGATGCGCGGCTCTCTATGAGGCTTGTGCGGTCGACAGCGGATCTGTACTGATGCTC 7440  
QY 7441 CTGTCCGAGCGGACGACGACCCCTGGAACCTGTGTCGAGCGGTTGCGGCGCGGACG 7500

Db 7441 CTGTCCGAGCGGACGACGACCCCTGGAACCTGTGTCGAGCGGTTGCGGCGCGGACG 7500  
QY 7501 GAGACCTGACGCACTCGGCGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7560  
Db 7501 GAGACCTGACGCACTCGGCGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7560  
QY 7561 GTGCTGCGGATCCGGGCTTCAACGAGCGCGCGGACGTCGTCGTCGTCGTCGTCGTCG 7620  
Db 7561 GTGCTGCGGATCCGGGCTTCAACGAGCGCGCGGACGTCGTCGTCGTCGTCGTCGTCG 7620  
QY 7621 GCCCACTGCGCGGCGGCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7680  
Db 7621 GCCCACTGCGCGGCGGCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7680  
QY 7681 GAGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7740  
Db 7681 GAGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7740  
QY 7741 ACCGAGCTGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7800  
Db 7741 ACCGAGCTGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7800  
QY 7801 GCGCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7860  
Db 7801 GCGCTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7860  
QY 7861 TCGGCGACCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7920  
Db 7861 TCGGCGACCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7920  
QY 7921 CAGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7980  
Db 7921 CAGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7980  
QY 7981 CTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8040  
Db 7981 CTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8040  
QY 8041 ACCGAGCTTCCGGAATGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8100  
Db 8041 ACCGAGCTTCCGGAATGAGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8100  
QY 8101 GTGACGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8160  
Db 8101 GTGACGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8160  
QY 8161 GACCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8220  
Db 8161 GACCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8220  
QY 8221 GCGGCGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8280  
Db 8221 GCGGCGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8280  
QY 8281 GCGGCGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8340  
Db 8281 GCGGCGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8340  
QY 8341 GTGCGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8400  
Db 8341 GTGCGACCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8400  
QY 8401 GTTCCGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8460  
Db 8401 GTTCCGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8460  
QY 8461 CCGTCGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8520  
Db 8461 CCGTCGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8520  
QY 8521 CCAATCTGTCGAGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 8580



Db 7487 GGTGAGCCCTGGCGGTACACGACGTGTGAAGCCGTGAGAGAAATGACTCTCGCCGCGGAC 7428  
Qy 2508 GCGGATCGATGCTGCGGCGCCGAGTGAAGTGTGGAAGACGACGATGGAAGTCCGGT 2567  
Db 7427 GGGGCTGACAGCCCTCTGTTCTCCAGCAGAAATCTCCACACCTCCGACAGGTGTGTCG 7368  
Qy 2568 CCGGGGTGACCCGCGCGCGAGTGAAGGAGTGGCCCCCGCGCAGCCCGGGGAAAACCGGCT 2627  
Db 7367 ACGGGGTGACCCGCGCGCGAGTGAAGGAGTGGCCCCCGCTCAGCTCCGGAGACCCGCT 7308  
Qy 2628 CCGGGTACAGGTGAAGTCTTCCAGAGAGTGTATCTGACCCGCACTGCGAATGGCGG 2687  
Db 7307 CCGGGTGAAGTGAAGTCTTCCAGAGAGTGTATCTGACCCGCACTGCGGAGGTCGCGG 7248  
Qy 2688 TCGGGGCAATGTTTCCGGCGCGATCCGCAACAGCTGGGGGTGACACCCCGACGAGGC 2747  
Db 7247 TGGAGCGCATGTCTCCGGCGGAGTGGCAGAGCTCGGCTCCACACCCCGCGCAGGC 7188  
Qy 2748 TCTCAACGCGTAACCAAGGTGGTCTGATGCTGACATGCCCCGGGTCGGCTCGGCGGAGTGA 2807  
Db 7187 TCTCAAGGGGTGACCCAGGTCAATGTGATGCTCCGGAGTCCGCTCGGCGGCGTATGTCGA 7128  
Qy 2808 CGAACCGGAGAGCCCTCCGCAACTCGGTCCGTTCCGCTCGAACACTGCGCTGCT 2867  
Db 7127 CGAACCGCACAGGCCCTCCGCAAGCTGTTGCGCTCGAGTCCGACAGCGCGCTGCT 7068  
Qy 2868 CCGGACCGCTGTAGTCTTCCGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2927  
Db 7067 CCGGCGCGCTGTAGTCTTCCGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 7008  
Qy 2928 CGTTGAGTTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2987  
Db 7007 CGTTGAGTCTGCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 6948  
Qy 2988 TGAATGTTGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 3047  
Db 6947 TGAATGTTGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 6888  
Qy 3048 CCGGAGAGTTCCTTCCGACCCGTTCCAGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 3107  
Db 6887 CCGGAGAGTTCCTTCCGACCCGTTCCAGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 6828  
Qy 3108 CGTGTGTTGTTCAAGCCGTTACAGGAGGTGAAGCAGCGCGGAGGTCCACAGGCGGAGT 3167  
Db 6827 CGTGTGTTGTTCAAGCCGTTACAGGAGGTGAAGCAGCGCGGAGGTCCACAGGCGGAGT 6768  
Qy 3168 GCGGCGGAGCGTCTTCTGTTGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 3227  
Db 6767 GCGGCGGAGCGTCTTCTGTTGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 6708  
Qy 3228 CCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 3287  
Db 6707 CCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGT 6648  
Qy 3288 CCGGAGAGGTGATCGCGGTGGGGTTCGCGACCGGACCTGTGTGAAGACGAGAGTGA 3347  
Db 6647 CCGGAGAGGTGATCGCGGTGGGGTTCGCGACCGGACCTGTGTGAAGAGCGGAGTGA 6588  
Qy 3348 GGAATCTGTTGCGGAGCACCGGACTCGCGGTGATCTTGGACCTGTGACCCGAGT 3407  
Db 6587 GGAATCTGTTGCGGAGCACCGGACTCGCGGTGATCTTGGACCTGTGACCCGAGT 6528  
Qy 3408 AGAAGTGAACGAGGAAATGACAGTGGGCGGAGGAGTGAAGCGCTGTGACGAGAGT 3467  
Db 6527 AGAAGTGAACGAGGAAATGACAGTGGGCGGAGGAGTGAAGCGCTGTGACGAGAGT 6468  
Qy 3468 CCGGCGTCCGCGCTACCGAGCGTCCAGCGCGCGCGCGGAGTGAAGCGCTGTGACGAGT 3527  
Db 6467 CAGGCTTGGAGTGAACCGCGGTTCAGTGGCGCTCGCGGTTCAGCGGTTCAGTGGT 6408  
Qy 3528 TGGTCAAGTACTTGGCGCGGCGGCTGTGACCGCGGTGACCGCGGTGACCGTGGAGTGCAT 3587

Db 6407 TGAACAGTACTTGGCGCGGCGGAGCGGCTCCAGCGGAGTCCAGGATCCGAGTGCAGC 6348  
Qy 3588 CGAACAGTTCAGACACTCGCCGAAACCCGCGCGGCTCAGGTGAACCGACGCGATCT 3647  
Db 6347 CGAACAGTTCAGACAGTGGCGAACTCCGCGGCTCAGGCGGAGCGGAGCGGCGATCT 6288  
Qy 3648 GCTTCAACGAGGAGTGAAGCGGAGGCGCGGAGTGGGCGGAGTGGGCGGAGCGGAGT 3707  
Db 6287 GCTTCAACGAGGAGTGAAGCGGAGTGGGCGGAGTGGGCGGAGTGGGCGGAGTGGG 6228  
Qy 3708 CTTGTGTTTCCGCGGTGCGAATGCGGACCGGTGAAGCGGAGTGAAGTGAAGTGAAGT 3767  
Db 6227 CTTGCAAGTCCGCGGTCAACCGTCAACCGCGCGGAGTGAAGCGGAGTGAAGTGAAGT 6168  
Qy 3768 CCGGACCAACCGCGCGCGGCGGAGTGGGCGGAGTGGGCGGAGTGGGCGGAGTGGG 3827  
Db 6167 CCGGACCAACCGCGCGCGGCGGAGTGGGCGGAGTGGGCGGAGTGGGCGGAGTGGG 6108  
Qy 3828 CCGCGGTGCGCGGAGGAGTGTCTTCAATGAGTGGGAGCCCTTCTGGGCTGTGTGTGTC 3887  
Db 6107 GGGCGGTTGCGGTGACAGAGTGTGTGATGGGAGGAGAACCTTCTGCACTTGTGTGTCGA 6048  
Qy 3888 GAGGATCAAGAACCGTTGCGTTTCCGTTGCTCCACTCCGCGTTGATCAACGACGCTG 3947  
Db 6047 GGTGTCAAGAACCGTTGCGGCTCTGTGTGCTCAATGGGCGTTGATCAAGCTTCCGCTG 5988  
Qy 3948 GTGTGCGGAGTGTGATGACGTGCAACCCGCGGAGTGTCTCACTGACACCGAGGG 4007  
Db 5987 GTGTGCGGAGTGTGATGACGTGCAACCCGCGGAGTGTCTCACTGACACCGAGGG 5928  
Qy 4008 CCGGTGCGGAGGCGGAGCCCGTTGGGAGGCACTCGGTCTGTGTGACGCGGAGCGGT 4067  
Db 5927 CCGGTGCGGAGGCGGAGCCCGTTGGGAGGCACTCGGTCTGTGTGACGCGGAGCGGT 5868  
Qy 4068 GGTCTCGTCCGAGTGAAGGCGGATCTGTGTGAGGCGGAGGAGGAGGAGGAGGAGGAGT 4127  
Db 5867 GGTCTCGTCCGAGTGAAGGCGGAGTGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGT 5808  
Qy 4128 ACCTTTCCGCGGAGGAGTGGCTGTGACGCGGATCAAGTGTGAGTGTGAGTGTGAGT 4187  
Db 5807 ACCTTTCCGCTCGGAGTGGCTGTGACGCGGATCAAGTGTGAGTGTGAGTGTGAGT 5748  
Qy 4188 TCCACAGGATGATCAAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4247  
Db 5747 TCCACAGGATGATCAAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5688  
Qy 4248 GTGACGCGCGGATACGTGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4307  
Db 5687 GATATCCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5628  
Qy 4308 TCCCTGTGTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4367  
Db 5627 ACCCTGTGTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5568  
Qy 4368 ATCTTGTGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4427  
Db 5567 ATCTTGTGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5508  
Qy 4428 ATGCGCGGATCGTTTCCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4487  
Db 5507 ATGCGCGGATCGTTTCCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5448  
Qy 4488 GCGGTGACCGCTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4547  
Db 5447 GAGGTGACCGCTTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5388  
Qy 4548 GGAAGTGAACCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 4607  
Db 5387 GGAAGTGAACCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5328  
Qy 4608 TGGTGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4667  
Db 5327 TGGTGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 5268



QY	4668	CCGACGATTCGCGCGGCTCTCGGGGTGACGAGCGCGCGGACGCTGATCCGGGTGACACGAA	4727
Db	5267	CCATGATCGGCGCGCTTCGGGTGTCAACGACGACGACGCTGCGGGGTGAGGAGG	5208
QY	4728	CCGATCGCCGCGGTTCGAGCTGCGACGAAACAACGGGTGTGAGTCCGAGCCAGCTCGCCGCTGC	4787
Db	5207	CCGGTCTCGGGGTTCAGTGGCGAACAACCGGTTCCAGCCCGACGCGACTCGCGCTGC	5148
QY	4788	GCGGTGGCCCGGAAGGTTCATTGACGCGATGATCATCTTCAACGGTGAAGTCAACCGGCCCCG	4847
Db	5147	GCGGTGGCCCGGAAGTTCATTGAGGAGATGACGACCTGCGCGGACAGGTGCTCGCGCGC	5088
QY	4848	AGCACGAGTTCCAGGGGCGACGGTGGCGTGTGAGGTGAGATGCAAGTCCCGTACCCGACG	4907
Db	5087	AGCACGAGTTGCGAGCGCGACCGTGGCGTTCAGAGTGGCCACGAGTGGCGGACACCCGCC	5028
QY	4908	AGGTGGCGGACCCGGGCGCTCGAATCTCCGCGACCGAGGGCGCGCGTTGTGAGCCAGTTG	4967
Db	5027	AGGTGGCGGACCCGGGCGCTCGAATCTCCGCGACCGAGTGTCCGCGTTGTGAGCCAGTTG	4968
QY	4968	TTGTTCAAGGGCCCATCTCAAGCGGGGCGAGGAAACGCTGCGCGGTGCGGATGCTGAGCGGG	5027
Db	4967	TTGTTCAAGCGCCCATCTCAAGCGGGGCGAGGAAACGCTTCGGGTCCCGAGCGTGGCGCTG	4908
QY	5028	CCGACGTGCGAGGGGGGTGCGAGGACGCTCGGGACCGCGGAGATCGCGAGATCGCATGCGTGGT	5087
Db	4907	CCGACGTGAGAGGGGTGTGCGAGGAAATGCTTCGGGGCGCGCGAAGATCGCGAGGTGCGTGCAGC	4848
QY	5088	ACGGCGCTTCAACCGCGGTGCGCGGTGTGTGAGATGCGCGGACGCGCAGCGGACGAGCTTGCGCA	5147
Db	4847	GCGGCTTTCACGCGCGGTCTCTGTGTGAGAGCGCGCGGCGCACGCGCATGTATGTGCGCA	4788
QY	5148	GCTGGATGTTCACGTAGGTGCTGTGCGCGGACGTTGCGTATCTGGCGGAGGTCAATCC	5207
Db	4787	GCTGGATGTTCACGTAGGTGCTGTGCGCGGACGTTGCGTATCTGGTCCAAAGTCAATCC	4728
QY	5208	ACCGGTGTGCGGGGCGACCGCGCTGCTGAAAGTCCGCGGGGCACTTCGACGAGCATGTATCC	5267
Db	4727	ACCGGTGTGCGGGGCGGACCTGCGCTGCGAAGTGGCGGGGCACTTCGATCAGCATGTATACC	4668
QY	5268	GCTTCTGTGTTGGGGTGAAGAACGCGCACCTCTCTGGAGTGCAGGCGGTGATGAGGAGATCT	5327
Db	4667	GCTTCTGTGTTGGGGTGAAGAACGCGCGCGCTCTCTGGAGTGCAGGCGGTGATGAGGAGATCT	4608
QY	5328	GGGTACCGGGGACGTCCAGCACGTAGTTCGAGGTAGGTTCGGCGGTGTTCTCGGGGACGT	5387
Db	4607	GCGAGCGCGGACCGTTCAACGATGATCAATATAGCCGGTTCGGAACTCTCGGGCGGT	4548
QY	5388	CGGTGTAGTGTTCGCGCTGACAGATGACACCGTTCGCGGCACTTCGGCGGACGTTCGGGCGAC	5447
Db	4547	CGGGTATTTGTTCGGGCTGGGACATGAGTGCACCGTAGGCGCGCATCTGGCGACGTTCATCCGC	4488
QY	5448	CCGCTCGGTTCGCGACGAGTGCAGACGCGGTGGAAGGTTCCCGTCAATCTCTTGACAGAGA	5507
Db	4487	CCGCTCGGTTCGCGACGATGTCAGAGCGGTGCACATGTCCCGCATATGTCCTTGACAGAGA	4428
QY	5508	GCGCGAGCATGCGCTGTGTTGGCGGGGGAGACGACGCGTTGCATCCAGGACTCACTTCGC	5567
Db	4427	GCGCGAGCATGCGCTGTGTTGGCGGGGCGAGACGAGCGCTGCATCCAGTGTGACCTTCGC	4368
QY	5568	GGTACGTGTGTGCTACGCGGACCGCGAAGATGTGAATATCTTGTGCTCGGTGCGA	5627
Db	4367	GGTGCTGTGTGCGCACCGTGAACCGGAAAGCTCGAATATCTTCTTCTCTGTGCTCGA	4308
QY	5628	TGCGGTGTGCGCGCGGATTCACCCGCTGCGGTGATGTGGGCAACGGCGCGGTGCGTT	5687
Db	4307	TGCGGTGTGCGCGCGGATTCACCCGCTGCGGTGATGTGGGCGAGCGGGGCGCGCGCT	4248
QY	5688	GAGCGAATCTCACTGTGTGTGCTGATATCCGAGATTCAGCTCAGCAGGGTGTTCATGTTGT	5747
Db	4247	GACGCAATCTGCGCTGCGCGGACGTCGAGAGATTCAGCTCAGCAGGGCTGTTCATGTTGT	4188

[illegible]

[illegible]

Db	7444	CTCCGGGGGTGAACCTCGAAACCGCCCGCCACCGCCAGCTCAGCTACTCGCAACTCCGGTTC	7385
Qy	2388	AGACGGCCGCGATGAGGCCCCGATGCTGATTCACACCCGTCGGCGACGTCGCCGTTCGAGTT	2447
Db	7384	ACCTTCCGCGCAGGAAGGCCCGCGGTGATCTCCACCCGTCGGCATGTCTGTTCGACT	7322
Qy	2448	GCTTGAGCCGGGCGGTCACCCGACTGTGTCTGAACCCGTGAGGAAGAACCTGTCTCCCGGCT	2507
Db	7324	GGTTGAGCCGGTCGGATCAACCACTGTGTGAAGCCGTGAGGAAGTACTCTGTCCGCGGGC	7265
Qy	2508	GCGATCGATGCTCCGCGCCGACGTTGACGAAGTGTGGAACAACGAGTGCAGGCTCCGGT	2567
Db	7264	GGGGCTGACGCGCTCTGTTCACGACAGAAATTCTCCACCACTCCGCGCAGCTGTGTG	7205
Qy	2568	CCGGGGGTGACCCCGCCCGCGCAGACTGAGCGGGTGGCCCCCGCCAGCCCGGGGAAACCGGCT	2627
Db	7204	ACGGGGGTGACCCCGCCCGCGCATGTGAGCGGTGGGCCCTCCAGCTCCGGGAAGCCCGCT	7145
Qy	2628	CCCGGTACAGGTGAACGTCTCCAGCAGGTGTGATCTGACCCGCACTTGCAGATGGCGG	2687
Db	7144	CGCGGTAGAGGTGACGTCCGCGCAGAGGTGTGACTTGCAACCGCCACCTGGGGGTGGCGG	7085
Qy	2688	TCCGGGCGATGTTCCGGCCCGGATCCGCGAACACTGTGGGCGTTCACACCCCGACGCAAGC	2747
Db	7084	TGGAGCGATGTCTCCGCGCGAGATGGCAGCACTGGGTTCACACCCCGCGCAGGC	7025
Qy	2748	TCTCCAAACGCTGAACCCGAGGTCCGTCTTGCAATGCCCCGGGTTCGCTCCGCGCGTAACTGA	2807
Db	7024	TCTCCAAAGGCGTAACTCCAGAGTCATGTGCAATGCCCGGGGTCCGCTCCGGCGCGTAACTGA	6965
Qy	2808	CGAAACCGGGCGAGGCCCTTCCCGCACTCCGTCGGTTCGCTCCGACAACTGTCGCTGT	2867
Db	6964	CGAAACCCGACCAAGGCCCTTCCGCGCAGCTCTGTGGCGCTCGAAGTCGCGCGCGCTGTGT	6905
Qy	2868	CCCGACCGCGTGTAACTCTCCCGGACCGGTGACGAAGTGAAGCGCGCGGTTCGGCGCTGA	2927
Db	6904	CGCGGCGCGTGTAACTCTCCGCGCACCGTCAAGAAAGTCAAGCGCGCTTTCGCGGCTGA	6845
Qy	2928	CGTTGAGTTCCGCGCATGAAGTCGACACAGGTCTGATGAGGCGGTTGGCCCGCCCGCGCAGA	2987
Db	6844	CGTTGACCTCCGCGCATGAAGTCGACAGAGTTCGATGAGCCGGTCCGCCCTGTCCCGCAGGA	6785
Qy	2988	TGATGTGTTGAGGCGCGAGCCGACCGGTGCGCGTCCGCTTCCGCGCGATCCGACAGAAGT	3047
Db	6784	TGATGTGTTGAGGCGCGAGCCGAGTCGGCGGTCTCCGCTCCGCGCGCATCCGACAGAAGC	6725
Qy	3048	CCCGGAGGTTCTTCTCTGACCCGTTCCAAAGCCGCAAGCTTTCGCGTGTGTGCGCTGTACT	3107
Db	6724	CTTGCAGAGTTCTTCTTAAACCGTTCCAAACCGCCGCTTTCGCGTGTGTGCTCTGTACT	6665
Qy	3108	CGTGTGTTCAAGCCCGGTACAGGGAGGTAGGAGCCCGGCGCGAAGTCCACACAGCCGGGCT	3167
Db	6664	CGTGTGTTCAAGCCCGGTAAAGGAGCGTGGAGATCCGCCCCAGCTCCACAGGCCGGGCT	6605
Qy	3168	GGCGGCGCAGCGTCTGTTCCGATGAGGCGGAAGCGTTGTGTAGAAGGTCAAGGCGGAAGC	3227
Db	6604	GGCGGTTCAAGGTCGTGCTCGGTGAGGCGGAAGGCGTTGTGTGTAGAAGGTTCGAAGAC	6545
Qy	3228	CCCGCCCGGGCGGTGCGACCAACGCCCCGATACCCCGGTTGTCAAGGTTTCAGGC	3287
Db	6544	CGCGCCCGGGCGGTGCGACCAAGCTCCGCGAGAACCGGGGTTGTGTACGGCGCTCGAGCC	6485
Qy	3288	CGCGGAGAGGTACATGCGCGTCCGGTTCTCCAGCGGCACTCGTCCGATGACGAGGTGA	3347
Db	6484	CGCGGACATGTACTGTGCGGTTGTCCGTGGGCACTCGTCCGATGATCCGGGCA	6425
Qy	3348	GGATCTGTTGCCGGGCAACCGACTCCGCGTGTGACTTTGGCACTGTCAACCCGTGACG	3407
Db	6424	GCGCTCTGTTGCCCGCGGAGACCGATCCGCGCTCTGTAGCGGGCAACCGGTACCCGCGACG	6365
Qy	3408	AGAAAGTACAGCGCGGAACATGCAAGTCCGGGCGCGGGGTGAGACCCGACGCTGTACGGGAAG	3467
Db	6364	AGAAAGTACAGCGGAACATGCAAGTCCGGGCGCGGGGTGACAGGCGGACGCTGTAGGGGAAG	6305





Db 4144 TGCCTGTCGCGCGGATCCACCCGCTCGATGTCGGGACGCGGCGCGCT 4085  
Qy 5688 GGAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5747  
Db 4084 GCAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4025  
Qy 5748 GGAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5807  
Db 4024 TGCCTGTCGCGCGGATCCACCCGCTCGATGTCGGGACGCGGCGCGCT 3965  
Qy 5808 GCAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5867  
Db 3964 GCAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3905  
Qy 5868 GCAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5927  
Db 3904 GCAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3845  
Qy 5928 GGAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5987  
Db 3844 GCAAGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3785  
Qy 5988 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6047  
Db 3784 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3725  
Qy 6048 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6107  
Db 3724 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3665  
Qy 6108 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6167  
Db 3664 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3605  
Qy 6168 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6227  
Db 3604 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3545  
Qy 6228 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6287  
Db 3544 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3485  
Qy 6288 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6347  
Db 3484 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3425  
Qy 6348 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6407  
Db 3424 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3365  
Qy 6408 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6467  
Db 3364 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3305  
Qy 6468 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6527  
Db 3304 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3245  
Qy 6528 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6587  
Db 3244 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3185  
Qy 6588 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6646  
Db 3184 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3130  
Qy 6647 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6667  
Db 3129 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3109

AAE30757  
ID AAE30757 standard; DNA; 47981 BP.  
XX  
AC AAE30757;  
XX  
DT 21-JUN-2001 (first entry)  
XX  
DE Micromonospora megalomicin biosynthetic gene cluster.  
XX Megalomicin; meg gene; polyketide synthase; antibiotic; motilide;  
KW antiparasitic; ds.  
XX  
OS Micromonospora megalomicina subsp. nigra.  
XX  
FH Location/Qualifiers  
FT complement(1..144)  
FT /tag= a  
FT /partial  
FT /gene="megBVI (megT)" /product="TDP-4-keto-6-  
FT /note="encodes AAB82201"  
FT 928..2061  
FT /tag= b  
FT /gene="megDVI"  
FT /product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
FT /note="encodes AAB82202"  
FT 2072..3382  
FT /tag= c  
FT /gene="megDI"  
FT /product="TDP-megosamine glycosyltransferase"  
FT /note="eryCIII homologue; encodes AAB802203"  
FT 3462..4634  
FT /tag= d  
FT /gene="megY"  
FT /product="mycarose O-acyltransferase"  
FT /note="encodes AAB82204"  
FT 4651..5775  
FT /tag= e  
FT /gene="megDII"  
FT /product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"  
FT /note="eryCI, DntV homologue; encodes AAB82205"  
FT 5822..6595  
FT /tag= f  
FT /gene="megDIII"  
FT /product="daunosaminyl-N,N-dimethyltransferase"  
FT /note="eryCVI homologue; encodes AAB82206"  
FT 6592..7197  
FT /tag= g  
FT /gene="megDIV"  
FT /product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"  
FT /note="eryVIII, dntV homologue; encodes AAB82207"  
FT 7220..8206  
FT /tag= h  
FT /gene="megDV"  
FT /product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
FT /note="eryIV, dntV homologue; encodes AAB82208"  
FT complement(8228..9220)  
FT /tag= i  
FT /gene="megBII-1 (megDVI)"  
FT /product="TDP-4-keto-6-deoxyhexose 2,3-reductase"  
FT /note="encodes AAB82209"  
FT complement(9226..10479)  
FT /tag= j  
FT /gene="megBV"  
FT /product="TDP-mycarose glycosyltransferase"  
FT /note="encodes AAB82210"  
FT complement(10483..11424)  
FT /tag= k  
FT /gene="megBIV"  
FT /product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
FT /note="encodes AAB82211"  
FT 12181..22821  
FT /tag= 1

```
FT /gene= "megaA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 1"
FT /note= "polyketide synthase; encodes AAB8212"
FT misc_feature
FT 12505..13470
FT /tag= m
FT /gene= "megaA"
FT /function= "AT-L"
FT 13576..13791
FT /tag= n
FT /gene= "megaA"
FT /function= "ACP-L"
FT 13849..15126
FT /tag= o
FT /gene= "megaA"
FT /function= "KS1"
FT 15427..16476
FT /tag= p
FT /gene= "megaA"
FT /function= "AT1"
FT 17155..17694
FT /tag= q
FT /gene= "megaA"
FT /function= "KR1"
FT 17947..18207
FT /tag= i
FT /gene= "megaA"
FT /function= "ACP1"
FT 18268..19548
FT /tag= b
FT /gene= "megaA"
FT /function= "KS2"
FT 19876..20910
FT /tag= t
FT /gene= "megaA"
FT /function= "AT2"
FT 21517..22053
FT /tag= u
FT /gene= "megaA"
FT /function= "KR2"
FT 22318..22575
FT /tag= v
FT /gene= "megaA"
FT /function= "ACP2"
FT 22867..23355
FT /tag= w
FT /gene= "megaA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 2"
FT /note= "polyketide synthase; encodes AAB8213"
FT 22957..24237
FT /tag= x
FT /gene= "megaA"
FT /function= "KS3"
FT 22957..24237
FT /tag= a1
FT /gene= "megaA"
FT /function= "KS5"
FT 24544..25581
FT /tag= y
FT /gene= "megaA"
FT /function= "AT3"
FT 24544..25581
FT /tag= aj
FT /gene= "megaA"
FT /function= "ATS"
FT 26230..26733
FT /tag= z
FT /gene= "megaA"
FT /function= "KR3 (inactive)"
FT 26230..26733
FT /tag= ak
FT /gene= "megaA"
FT /function= "KR5"
FT 26998..27258
FT misc_feature
```

```
FT /tag= aa
FT /gene= "megaA"
FT /function= "ACP3"
FT 26998..27258
FT /tag= a1
FT /gene= "megaA"
FT /function= "ACP5"
FT 27393..28590
FT /tag= ab
FT /gene= "megaA"
FT /function= "KS4"
FT 27393..28590
FT /tag= am
FT /gene= "megaA"
FT /function= "KS6"
FT 28897..29931
FT /tag= ac
FT /gene= "megaA"
FT /function= "AT4"
FT 28897..29931
FT /tag= an
FT /gene= "megaA"
FT /function= "AT6"
FT 29953..30477
FT /tag= ad
FT /gene= "megaA"
FT /function= "DH4"
FT 29953..30477
FT /tag= ae
FT /gene= "megaA"
FT /function= "RR4"
FT 31396..32244
FT /tag= ae
FT /gene= "megaA"
FT /function= "RR4"
FT 31396..32244
FT /tag= ap
FT /gene= "megaA"
FT /function= "ACP6"
FT 32257..32799
FT /tag= af
FT /gene= "megaA"
FT /function= "RR4"
FT 32257..32799
FT /tag= ag
FT /gene= "megaA"
FT /function= "ACP4"
FT 33052..33312
FT /tag= ag
FT /gene= "megaA"
FT /function= "ACP4"
FT 33666..43271
FT /tag= ah
FT /gene= "megaA"
FT /product= "megalomicin 6-deoxyerythronolide B synthase 3"
FT
```

Query Match 29.2%; Score 2606.8; DB 4; Length 47981;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 6415 CTCGAGCCGATGCTGCGCGCGCGGTGGGCCAACAGTCGTGACGTCGTCGTCGCGGT 6474
DB 1 CTCGAGCCGATGCTGCGCGCGCGGTGGGCCAACAGTCGTGACGTCGTCGTCGCGGT 60
QY 6475 GGGAGGTCGCGCGCGCGCGCGAGTCAGAGAAACGTAATGCGATTGCGATTCCGAGTCGC 6534
DB 61 GGGAGGTCGCGCGCGCGCGAGTCAGAGAAACGTAATGCGATTGCGATTCCGAGTCGC 120
QY 6535 ATGACCGTTGACCCGATCCCGCATACGCTCTCCCGTGAATGTCGTGGCGGTCGTCGCG 6594
DB 121 ATGACCGTTGACCCGATCCCGCATACGCTCTCCCGTGAATGTCGTGGCGGTCGTCGCG 180
QY 6595 TACCGCCCGACTGACATTCGTCGATCAAGACCCCGCCAGTGTAGGAGCTCCGCCCGCA 6654
```

Db	181	TAACCCCGAGCTGAATTCGTGATCAAGAACCCGCCAGTGAAGGCTCCGCCGCA	240
OY	6655	CGGAGGAAGTCCGTGGAACAATTCGCGGTGACCGGTCCGCCGTCGTGAACAAGGAC	6714
Db	241	CGGAGGAAGTCCGTGGAACAATTCGCGGTGACCGGTCCGCCGTCGTGAACAAGGAC	300
OY	6715	GTCGAGCAACCCGATCATTTGCTGTCCGTGAACCTTCCTAACTGTCCGCGGACATCTTTC	6774
Db	301	GTCCGAGCAACCCGATCATTTGCTGTCCGTGAACCTTCCTAACTGTCCGCGGACATCTTTC	360
OY	6775	TGAACCGGTGTGTTCCGTGATGAAGCGGTTCCCGGCCGCTGTGAACCTGTGCTGGACT	6833
Db	361	TGAACCGGTGTGTTCCGTGATGAAGCGGTTCCCGGCCGCTGTGTGAACCTGTGCTGGACT	420
OY	6835	GACCGGTTGCCGCGTGTGTTTCGCCCGCTTTCGAACTGCGGATTCGTGATCGCGGAGTG	6894
Db	421	GACCGGTTGCCGCGTGTGTTTCGCCCGCTTTCGAACTGCGGATTCGTGATCGCGGAGTG	480
OY	6895	GGAGCGGGTGTCTGAACCGGATGATCTGCAATCATGGCGCTCAATGAAGATCTTGTAG	6953
Db	481	GGAGCGGGTGTCTGAACCGGATGATCTGCAATCATGGCGCTCAATGAAGATCTTGTAG	540
OY	6955	CATGGTCCGCGCCGAGGGTCCGACAGGCCCGGAACGCCCGGATCAAGCTGTGTGACGA	7014
Db	541	CATGGTCCGCGCCGAGGGTCCGACAGGCCCGGAACGCCCGGATCAAGCTGTGTGACGA	600
OY	7015	CGTCGACATCAACCGTGAAGCCGCGATGACCCGACACAGCGCATGCTGGTCCGCACT	7074
Db	601	CGTCGACATCAACCGTGAAGCCGCGATGACCCGACACAGCGCATGCTGGTCCGCACT	660
OY	7075	GGAAGGATGAGCGCATCAAGGAAATGAGCGTGTCACTAGACAGAGCGCAACAGCTGTCC	7134
Db	661	GGAAGGATGAGCGCATCAAGGAAATGAGCGTGTCACTAGACAGAGCGCAACAGCTGTCC	720
OY	7135	GGGCTGTGGGAAACAGCATTCGATCTGCGTCAAGCGTTCAATTGCCCGCGGACCGCTT	7194
Db	721	GGGCTGTGGGAAACAGCATTCGATCTGCGTCAAGCGTTCAATTGCCCGCGGACCGCTT	780
OY	7195	GGAATTCGATGACCGGTCCGTCCGACATGACGATCGGGAACCGCGGTTTGAGACACA	7253
Db	781	GGAATTCGATGACCGGTCCGTCCGACATGACGATCGGGAACCGCGGTTTGAGACACA	840
OY	7255	GGTAGTAGCGATGACGCGTGTCTCTCCGCGCGGACGCGTCCGACATGAGTGAATCCG	7314
Db	841	GGTAGTAGCGATGACGCGTGTCTCTCCGCGCGGACGCGTCCGACATGAGTGAATCCG	900
OY	7315	TCAAGTCTTCAATCCGGAACGTTCTAATGCGATGTGCGATCGAAGCGGCTGGACCGG	7374
Db	901	TCAAGTCTTCAATCCGGAACGTTCTAATGCGATGTGCGATCGAAGCGGCTGGACCGG	960
OY	7375	GAGTTGCAAGATGCGCCCGGGGTCTCACTGGGGGTTCCGTGCAATGGCGATCTGTACTCG	7434
Db	961	GAGTTGCAAGATGCGCCCGGGGTCTCACTGGGGGTTCCGTGCAATGGCGATCTGTACTCG	1020
OY	7435	ATGCTCCGTGTCGAGCGGGAAGACACCCCTGTGAACCTGTGAAGAAAGGTTGCGGGCGGC	7494
Db	1021	ATGCTCCGTGTCGAGCGGGAAGACACCCCTGTGAACCTGTGAAGAAAGGTTGCGGGCGGC	1080
OY	7495	GGACGGGGAACCGTACCGCCAGTCCGAGCGGAAAGTGGAGTTCGTGTACCAACCGACCGC	7554
Db	1081	GGACGGGGAACCGTACCGCCAGTCCGAGCGGAAAGTGGAGTTCGTGTACCAACCGACCGC	1140
OY	7555	GCCAAGGTGTCTCCGCGAATCCCGGGTTCAACCAACCGGCCCGCGCAAGCTGCCGTGTGATG	7614
Db	1141	GCCAAGGTGTCTCCGCGAATCCCGGGTTCAACCAACCGGCCCGCGCAAGCTGCCGTGTGATG	1200
OY	7615	CAGGTGATCACTGGCCGCGGAGCTCTCTGAGGCGGAGCCCTTCGAGGATTTCAAGCCCGC	7674
Db	1201	CAGGTGATCACTGGCCGCGGAGCTCTCTGAGGCGGAGCCCTTCGAGGATTTCAAGCCCGC	1260
OY	7675	ACCGAGACGCGCGCTGTCTGTGACAGATGAGCGCTCAATGGCTTCAGACAGGTCGCGCAGG	7734

Db	1261	ACCGAGACGCGGCGCTCGGTGACATGTGACAGCCGACCTGCGTCCAGACAGCGCTGCGCCAGG	1320
QY	7735	CTGTGTACCCAGACTGGGGGTGCGGCTTCGATCTGTGTAAACAATTGCGCCGGAGAGTCCGG	7794
Db	1321	CTGTGTACCCAGACTGGGGGTGCGGCTTCGATCTGTGTAAACAATTGCGCCGGAGAGTCCGG	1380
QY	7795	GTGCTTGGCGCTCGGTATCGCGGCCCGGCACTCAAGGGGTGTGACCCCGACCGTCTTCGAGTCC	7854
Db	1381	GTGTGTGGCGCTCGGTATCGCGGCCCGGCACTCAAGGGGTGTGACCCCGACCGTCTTCGAGTCC	1440
QY	7855	TGGAACTGTGGCAACCCCGGTATAGCTCTGTGACGCTCCAGAGTCAGCCCCGACACAGCTTGCAGTG	7914
Db	1441	TGGAACTGTGGCAACCCCGGTATAGCTCTGTGACGCTCCAGAGTCAGCCCCGACACAGCTTGCAGTG	1500
QY	7915	ACCGAACAAGGCGCTGACCGGCGCTTCGACGAGATGTGACGCGGATGTACCGAGCGGTCCGGAAGCC	7974
Db	1501	ACCGAACAAGGCGCTGACCGGCGCTTCGACGAGATGTGACGCGGATGTACCGAGCGGTCCGGAAGCC	1560
QY	7975	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	8034
Db	1561	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	1620
QY	8035	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	8094
Db	1621	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	1680
QY	8095	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	8154
Db	1681	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	1740
QY	8155	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	8214
Db	1741	GCGGTGTGTGTGGGGGTGTGTGGCGGAGCTGTGGCGGCAACAACGTGTGGCAACGCGCTTCTG	1800
QY	8215	GTGCGCGCGGCGAACCTGTGATCCGAGAGTCTTCAACCGATCCGACCGGATTCCAGCTGTGAC	8274
Db	1801	GTGCGCGCGGCGAACCTGTGATCCGAGAGTCTTCAACCGATCCGACCGGATTCCAGCTGTGAC	1860
QY	8275	GCTTGGCGGCGAACCTGTGATCCGAGAGTCTTCAACCGATCCGACCGGATTCCAGCTGTGAC	8334
Db	1861	GCTTGGCGGCGAACCTGTGATCCGAGAGTCTTCAACCGATCCGACCGGATTCCAGCTGTGAC	1920
QY	8335	GCGCTGTGTGACCACTGTGACCACTGTGACCACTGTGACCACTGTGACCACTGTGACCACTGTGAC	8394
Db	1921	GCGCTGTGTGACCACTGTGACCACTGTGACCACTGTGACCACTGTGACCACTGTGACCACTGTGAC	1980
QY	8395	CTGTCCGCTTCCGGGCGCGGTGATCAAGACGATCGGTGACCCGTCGCTCGCTCGCTGATCAAG	8454
Db	1981	CTGTCCGCTTCCGGGCGCGGTGATCAAGACGATCGGTGACCCGTCGCTCGCTCGCTGATCAAG	2040
QY	8455	CGTTGGCGGCGGTGATCAAGACGATCGGTGACCCGTCGCTCGCTCGCTGATCAAGACGATCGGTGAC	8514
Db	2041	CGTTGGCGGCGGTGATCAAGACGATCGGTGACCCGTCGCTCGCTCGCTGATCAAGACGATCGGTGAC	2100
QY	8515	CAACAGCGCATCTGTTCGCGGTGATCAAGACGATCGGTGACCCGTCGCTCGCTGATCAAGACGATCGGTGAC	8574
Db	2101	CAACAGCGCATCTGTTCGCGGTGATCAAGACGATCGGTGACCCGTCGCTCGCTGATCAAGACGATCGGTGAC	2160
QY	8575	GATACGAGGTGTGTGCGCTGTGCGCGGCGCTGACCGACGACGATGTCAACCGATGTCTGTGACCGG	8634
Db	2161	GATACGAGGTGTGTGCGCTGTGCGCGGCGCTGACCGACGACGATGTCAACCGATGTCTGTGACCGG	2220
QY	8635	CGTGGCGCTCGGTGACCACTGTGAACTTGTGTGAGTGTGACCGCCACGCGGCGCGGACGAT	8694
Db	2221	CGTGGCGCTCGGTGACCACTGTGAACTTGTGTGAGTGTGACCGCCACGCGGCGCGGACGAT	2280
QY	8695	CGTGGAGTACATGCGGACCTCTGATCGGTGACCAAGACCAACAATCTCTGTGAG	8754
Db	2281	CGTGGAGTACATGCGGACCTCTGATCGGTGACCAAGACCAACAATCTCTGTGAG	2340
QY	8755	CGACCTCTGTGGAGTACAGACCACTTGTCAACCCCGACCTTCTTGTGCTGTGATGACCCCGA	8814
Db	2341	CGACCTCTGTGGAGTACAGACCACTTGTCAACCCCGACCTTCTTGTGCTGTGATGACCCCGA	2400

QY 8815 CTCGCTCATGACGAGGATGTCAGATTCTGCGCTCTGCGGTCGCCAGCTGATCGTCTG 8874  
DB 2401 CTCGCTCATGACGAGGATGTCAGATTCTGCGCTCTGCGGTCGCCAGCTGATCGTCTG 2460  
QY 8875 GAGCGCGCTGACCTTCG 8934  
DB 2461 GAGCGCGCTGACCTTCG 2520  
QY 8935 GATGCTGTGGGGTCCGAGCGTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8994  
DB 2521 GATGCTGTGGGGTCCGAGCGTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2580  
QY 8995 CCACCGAGAGGTGAGGACCGCGAGATCC 9024  
DB 2581 CCACCGAGAGGTGAGGACCGCGAGATCC 2610  
RESULT 5  
AA287284  
ID AA287284 standard; DNA; 12441 BP.  
AC AA287284;  
AT 15-SBP-2003 (revised)  
DT 05-JUN-2000 (first entry)  
XX  
DB 5. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.  
XX  
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;  
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;  
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;  
KW chronic obstructive pulmonary disease; respiratory inflammation;  
KW hypercholesterolaemia; crop protection agent; ds.  
OS Streptomyces venezuelae; ATCC15439.  
PN MO20000620-A2.  
PD 06-JAN-2000.  
XX  
PF 25-JUN-1999; 99WO-US014398.  
XX  
PR 26-JUN-1998; 98US-00105537.  
XX  
PA (MINU ) UNIV MINNESOTA.  
PI Sherman DH, Liu H, Xue Y, Zhao L;  
PI WPI; 2000-160679/14.  
DR P-PSDB; AAY77179.  
XX  
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
PT synthesis of methymycin and pikromycin.  
PS  
XX Claim 2; Page 281-287; 438pp; English.  
CC The invention relates to an isolated and purified nucleic acid segment  
CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
CC biologically active variant, where the nucleic acid sequence is not  
CC derived from the erylC gene cluster of *Saccharopolyspora erythraea* or  
CC streptomycetes antibiotics. The invention also relates to a macrolide  
CC biosynthetic gene cluster, or fragments thereof. The macrolide  
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,  
CC pikromycin, neomethymycin, narbomycin or a combination of these  
CC compounds. Recombinant or augmented cells comprising the desosamine  
CC and/or macrolide biosynthetic gene clusters are useful for the production  
CC of biologically active macroclides. The macrolide biosynthetic proteins  
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and  
CC narbomycin. The alternative termination of polyketide synthesis may be  
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
CC monomers. The compounds produced by the recombinant host cells are useful  
CC as biopolymers, e.g., in packaging or biomedical applications, to

CC engineer PHA monomer synthases or to prepare biologically active agents,  
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
CC chronic obstructive pulmonary disease as well as other diseases involving  
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based  
CC antibiotics which are active against a variety of organisms, e.g.,  
CC bacteria, including multi-drug resistant pneumococci and other  
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
CC protection agents (e.g., fungicides or insecticides) via expression of  
CC polyketide in plants. The present sequence represents the desosamine  
CC biosynthetic gene cluster from *Streptomyces venezuelae* ATCC 15439.  
CC (Updated on 15-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 U; 0 Other;  
Query Match 14.6%; Score 1309.4; DB 3; Length 12441;  
Best Local Similarity 67.7%; Pred. No. 3.9e-170;  
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;  
QY 2369 GGGCGGGTGCAGCGCGTTCAGACGCGCGGATCAGCGCGCGCGCGCGCGCGCGCG 2428  
DB 9472 GGGCGGATCGGGGCGGGTGCAGCGCGGATCAGCGCGCGCGCGCGCGCGCGCGCGCG 9531  
QY 2429 GCGACGTCCGCTTGCAGATTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2488  
DB 9532 GCGCGGTGCGCTTCAGACTGATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9591  
QY 2489 AAGAACTGTCCCGCGCTCGGATCGATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2548  
DB 9592 AAGTATCTGTGCGCTTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9651  
QY 2549 ACCGAGTGCAGGCTCCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2608  
DB 9652 ACCTCGGTGAGGAGGTGTGCGGGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9711  
QY 2609 AGCCCGGGGAAACCGCGCGTCCCGGTACAGGTGACGTCTCCGACAGGTGATCTGACAC 2668  
DB 9712 AGGTGAGGAAAGCGCGCTTGCAGTACAGGTGACGTCTCCGACAGGTGATCTGACAC 9771  
QY 2669 GCCACTGCGGATGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2728  
DB 9772 GCGACCTGCGGGGTGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9831  
QY 2729 TCGACACCGCGAGCGAGCTTCTCCAGCGGTGACCGAGGTGCTGCGATGCCGGGGTC 2788  
DB 9832 TCGGCGCGCGGTGCGAGGCTGTTCAGGCGCGGTGACGTGATGATGATGATGATGATG 9891  
QY 2789 CGCTCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2848  
DB 9892 CGCTCGCGGAAACCGCGCTCTCGAAGCGGTGAGGGCGCTCTGAGAGCTCGCGCGCTCC 9951  
QY 2849 TCGGACACCTGCGCGGTGCTCCGACCGCTGATGCTCGCGAGCGGTGACGAAATCGAGC 2908  
DB 9952 TCGGCGAGCTTGCCTGTGATGACCGCGCGCTGATGCTCGCGAATGTTGACAAATCGATC 10011  
QY 2909 GCGCGGTGCGGCTGAGCTCGTTGATGTTGCGCGATGATGATGATGATGATGATGATGATG 2968  
DB 10012 GTCTGCGCGTCCCGCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 10071  
QY 2969 TTGCGCGCGCGCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028  
DB 10072 GAGGACAGCGCGCGGAGACATATGAGCGAGGCGAGTTGATGCGGATGCGCGCTCG 10131  
QY 3029 GCGCGCATTCGCGAGAAAGTCCCGAGGTTCTTCTGACCGCGTTCGAAAGCGCGCATTTG 3088  
DB 10132 GCGCGAGACTGTGAAAGCGCGAGGTTCTTCTGCGAGCGCGCGGAGGCGCGCTTTTG 10191  
QY 3089 CCGGTGTGCGCTGTGATCTGTCTGTTTCAAGCCCGTACAGAGAGTACGAGCGCGCTG 3148  
DB 10192 CCGGTGTGCTGTGATCTGTCTGTTTCAAGCGCGTGTGAGCGAGGTGCGATGCGCGTGC 10251  
QY 3149 AGGTTCACAGAGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3208  
DB 10252 AGGCGCGAGAGCGCGGTGCGCGCTTCAGAGGTGCGCTCGGTGAGCGCGAGAGGATTCGTG 10311

OY	3 269	TAAGAAGGTCAAGAGGAGAAAGCCCGCCGGGCGAGCTGTGCACAACCAAGCCCTCGATATCCCGG	3 268
Db	10312	TAGAAGGTGAGGAGCGCAGAGCGGTGGTTCGTGTGTCGTGTGCGAGCGCTCCGAGCGCGAG	10371
OY	3 269	TTGGATCAACGGATTTCAGAGCGCGCGGAGAGGATATCGCCGGTCGGATTCTCCAGCGGACCG	3 328
Db	10372	TTGGATTGACGGAGCTCAAGCGCGCCCGGAAAGTATATGCGCCAGAGGGATTGCCCGGGATTATC	10431
OY	3 329	TCCGTGATGACCGAGGTGAGGATCTCGTTTGCACCGGAGCACACCGACTCCGCGTGTATCTTG	3 388
Db	10432	TCCGTGATGACCGAGCCGAGAACATATGCGCTTTCGCGCGTTCAGAGGGCGGACGGGTGTATGCGG	10491
OY	3 389	GCACCTGTGCACCCGTAACGCGAAGATGACAGCGGAAATATGCAACTGTCCGACCGGGATGAGAC	3 448
Db	10492	GCGCCGGGTCAACCGGACCGAAGAGTGTGCACCGGAACTATGCAAGGTCCGAGCCGAGGGTATGAG	10551
OY	3 449	CCGACGCGGTATCCGGAAGAGGAGGCTCCGCGGTATCCGCGAGCGCTCAACCGCGCCGAGGCGG	3 508
Db	10552	CCGACGCGGTATCCGGAAGAGGAGGCTTCGTGCGAGCGCGCGCTGTGAAGACGCGCGCTGT	10611
OY	3 509	TTCAAGCGGCTGATCTGTGTGTCTCAATATCTTTCGCGCGGAGCGCGCTTCAACCGCGGTG	3 568
Db	10612	TTCAGCGGAGACAGGGGTCTTTCAGATATAGCCCGCGGGGCGGGCTTCGACCGCGGGT	10671
OY	3 569	CCAGCGTTCGAGGATCGATCGAACAGGGTTCGAGCACTCCGCGAACCCGAGCCCGGTTCAGG	3 628
Db	10672	CCGAGCTTCGGGAACTTCCCGGAACAGGGCGAGGAGGCGCGAAGGCGTCCCGGTGACG	10731
OY	3 629	TTCGAACCGAGCGGCAATCTGTCTCAATCGGGGTGAACGGCGAGGCGCGGTATGTAGTGGGTG	3 688
Db	10732	CCGAGGTCTGTGCGAGGCTCTCTCGAGCGGGGTAAAGGGGTGTGTGCGTATAGCGCACAGGCG	10791
OY	3 689	AGTTCCGCGAGACCGGACCGGCGTGTCTGTTCGCGCGGTGTCCGAGTCCGAGACCGGTAAGGCGG	3 748
Db	10792	AGCCGGAACGAGGTATCGGGGCGGGTCTTCGAGCTCTGTCCGAGCGGACAGAGCCGCGGGCG	10851
OY	3 749	GTGAACCTTCGGCGGTGAAGCGCCCGCAACCAACGAGCGGAGCGGTCGAGTCTGTGCAGCG	3 808
Db	10852	GCGAGGGGTCTGAGCGGAACGGCGTGAACGCGCGCCCGCAATCCGCTCCGGGGTGTGGCAGG	10911
OY	3 809	TCCGCGATCTCCGTCCGTATAGCGGGGTTCGCGTATGAGTGGTCTTCAATCGACGTGCGAAC	3 868
Db	10912	CGTTTCGAGCGGAGCGGTGCGAAGAGGAGCGGAGCGGTCAATCGGAGCGTCAATGTAGG	10971
OY	3 869	CTTTCGCGGTCTGTGTGTGCGAGGATCAACGAACGTTTCCGTTTCCGCTTCCCATCTCCGCG	3 928
Db	10972	GCGGTGATGTCTGAGGAGGAGCGCAGAGCGGAGCGGAGCGCGGTCTGTCCGAGTGC-----GCGC	11026
OY	3 929	GTTCGATCAACGCAACCGCTGTGTGTGCGAGGTCCGATATGACGTTCGCAACCCCGGCGGATGTC	3 988
Db	11027	GTTCGATGTTCCGCGGCGCGGGGTCCGCGAGAGAACCGACAGGTTCGAGCAACCCGCGGATGTC	11086
OY	3 989	CTCACTGACAACCGAGGAGGCGGAGTCGGAAGGAGGACACCCGTTCCGCGAGCACTCCGT	4 048
Db	11087	GTTCGTCGCGGATGAGCGGTGCGGAGTCGCGGAGGAGGACACACGCGCGCGGAGGCTTCGCT	11146
OY	4 049	CTGTGTGAGCGCGACGCGGTGTCCGTCCGTATAGGCGGTCAATTTGGTGTGCAGCGCGGAG	4 108
Db	11147	GTGTGAGGAGCGGAGCGGTGCGGAGTCCCGCGGTATCGGCTCAAGTCTGTGGCAACCCGCGCA	11206
OY	4 109	GAAATAGAGGCTGTGTGCGACGACTTCTTCGCGCGCGAGATTCGCTTCGACGCGGTCACGTC	4 168
Db	11207	GAAATAGGCGCGGAGTGTGCACGCGCTTTCGCGCTTCAAGAACCTTCATATAGAGGTTCGCGTG	11266
OY	4 169	GATTCGCGGTGCGGAGTCCGATCGAACAGATGATCAAGTATCTGTGTATTTGCTCTCTCTGTC	4 228
Db	11267	GATTCGCGGTGAGGCTCTGTGATCTCGAGATCACTATCTGTGTGTTGTATGAGGCGGTG	11326
OY	4 229	GAGCGGAGAGGATGACAGGTGACAGGTGACGCGGTACTGTCGCGCACTGCTGTGTATGAGCGC	4 288
Db	11327	GCGGTCTGTGTCCGCGACAGAGACGCGCGGAGAGGTTCGCGAGGTGTCTCCGCTATAGCGCGC	11386

QY	4289	TTGGTTGACCCGGTGTGTGCTCCCTCGTCTTGCGCGAAGCGCTTCAGAGGAGGTGAGCCCAT	4348
Db	11387	GTGGTTGGCGCGGTTCGGCTCGAATGAACCTTGGGAAAAGCGTTCGAGGGAGGTGAGGCCAT	1144
QY	4349	GGCGCGGGCGCACTCGCTCATCTTGCGGTTGTGTGTCGCATCTCGGTGAACA	4408
Db	11447	GGCGCGGGCGGCGCTTCGCTCATCTTGGGTTGTGTCCCGCGGCGGGCTGCGCGCGGCGAG	1150
QY	4409	GCAGATGCGAAGTTGTGTCATGCGCCCGAATCCGTTCCGCGCAGCGAGCGCGTGTGATGAC	4468
Db	11507	GTCCGAAGCGGAAGTTGTGAGAGGGCGCGGAATCCGGCGGGCGAGAGTCCGCGGTGTGATGAC	1156
QY	4469	CACGCGCCCGCGCTCTGAAAGCGGTGACCGCGCTTGTGTGTGCGGTGAAGCTTGAACA	4528
Db	11567	GACGCGCGCGCGCTCTGAAAGGGTGTGAAGCGCTTGTGTGTGTGTGAAGCTTGAAGACTTCGCGC	11622
QY	4529	GTCAACCGGATCCGCGCCACCGGACGATCCACCGTGTGTACACCCAGAGGGCGTGGCGCGTC	4588
Db	11627	GTCCGCGAAGCTTCCGCGCGGCGCGCGCGCTGTGACCGCGCAACCGAGGCGGTGCGCGGCGTC	11688
QY	4589	GAAAGAAGCTTGAACCTGTGTGTGTGCGCGCGGATCTTTCGACGCGCGCTCCACAGATGCTGG	4648
Db	11687	GAAATGACAGCGCGCGAGCGCGGTGTGTGTGCGGACCTTTCGAGAGCTGTGTGCGCGCGACAGG	11744
QY	4649	TGCGCCCCACAGGTGTACGCGCGACGATCGCGCGCGGTCTGCGGGGTGAAGAGCGCGAC	4708
Db	11747	GCGCGCCCCAGAGGTGTGACCGCGCAAGCGCGCGAGTGTGCGGGGTGTGACCGCGCGCGCAC	11800
QY	4709	GTGGTCCGCGGTGACCAAGACCGGTCGCGCGGCTGTGACGTGTGACAAACAACCGGTGTGATCC	4768
Db	11807	CTGGTCCGCGGTGTGAGTTGGCCGCGTGTCCGCGGTGTGATGTGTGCGCAAGACTCGGGGTGAAGCC	11866
QY	4769	GAGCCAGCTCGCGCGGTGTGCGCGCGGTGTGCGCGAAGGTCAATTGACGACATGATCACTTCACC	4828
Db	11867	GATCCAGCGCAGT	11922
QY	4829	GGTGACGTCAACCGCGCGCGACCAAGTTCAGGCGCGAGCGGTGTGTGTGTGTGTGTGTGTGTGT	4888
Db	11927	GGTGAAGCG	11986
QY	4889	GCAATGCGCTTACCCCGACCAAGT	4948
Db	11987	GGCATGCG	12044
QY	4949	GCCGTTGGTGTGAGCAGT	5008
Db	12047	GCCGTTGTGACAGCAGCTGT	12100
QY	5009	GTCGCGCATGT	5068
Db	12107	GTGTGATGT	12166
QY	5069	GATGCGCAGATGT	5099
Db	12167	GAATGCGAGGT	12197
RESULT 6			
ID	AAD39043	standard; DNA, 13613 BP.	
AC	AAD39043;		
DT	23-SEP-2002	(first entry)	
DS	Streptomyces venezuelae desosamine gene cluster.		
KW	Glycosylated polyketide; modified recombinant bacterial host cell; mRBHC;		
KW	macrolide; anthracycline; angucycline; avermectin; milbemycin;		
KW	tetracycline; polyene; polyether; ansamycin; isochromanone; sugar;		
KW	desosamine; des; gene; ds.		
OS	Streptomyces venezuelae.		



XX	Key	Location/Qualifiers
FT	CDS	3. .809
FT		/*tag= a
FT		/product= "Streptomyces venezuelae Des gene cluster
FT		encoded protein #1"
FT		806. .2014
FT		/*tag= b
FT		/product= "Streptomyces venezuelae Des VIII protein"
FT		/transl_except= (pos:806. .808, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	2162. .3442
FT		/*tag= c
FT		/product= "Streptomyces venezuelae Des VII protein"
FT		3535. .4245
FT		/*tag= d
FT		/product= "Streptomyces venezuelae Des VI protein"
FT		/transl_except= (pos:3532. .3534, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT		4312. .6741
FT	CDS	/*tag= e
FT		/product= "Streptomyces venezuelae Des R protein"
FT		/transl_except= (pos:4312. .4314, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(7969. .6829)
FT		/*tag= f
FT		/product= "Streptomyces venezuelae Des V protein"
FT		/transl_except= (pos:7969. .7967, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(8979. .7966)
FT		/*tag= g
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		/transl_except= (pos:8979. .8977, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
FT	CDS	complement(9111. .9989)
FT		/*tag= h
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		complement(11639. .10182)
FT		/*tag= i
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		complement(12883. .11636)
FT	CDS	/*tag= j
FT		/product= "Streptomyces venezuelae Des IV protein"
FT		/transl_except= (pos:12881. .12883, aa:Met)
FT		/note= "CDS does not include start codon"
FT		/partial
XX	PN	MO200229035-A2.
XX	PD	11-APR-2002.
XX	PE	05-OCT-2001; 2001WC-US031255.
XX	PR	05-OCT-2000; 2000US-0238185P.
XX	PA	(MINU ) UNIV MINNESOTA.
XX	PA	(LIUH/) LIU H.
XX	PA	(SHSR/) SHERMAN D H.
XX	PA	(ZHAO/) ZHAO L.
XX	PI	Liu H, Sherman DH, Zhao L;
XX		WPI; 2002-405171/43.
XX		P-PSDB; AAE24228, AAE24229, AAE24230, AAE24231, AAE24232, AAE24233,
XX		AAE24234, AAE24235, AAE24236, AAE24237, AAE24347.
PT		Modified recombinant bacterial host cells in which the expression and
PT		activity of nucleic acids encoding sugar biosynthetic enzymes has been

altered, useful for producing metabolites with altered sugar structures.

Disclosure; Fig 8; 174pp; English.

The invention provides a method to alter the sugar structure diversity for a particular metabolite via the recruitment and collaborative action of sugar genes from a variety of sugar biosynthetic pathways to yield a metabolite comprising a non-natural sugar, e.g.; a novel glycosylated polyketide. The invention also relates to a modified recombinant bacterial host cell (mBHC) in which the expression and activity of nucleic acids encoding sugar biosynthetic enzymes has been altered. The mBHCs may be cultured to produce the modified sugar products, e.g. a macrolide, anthracycline, amphotericin, avermectin, milbemycin, tetracycline, polyene, polyether, ansamycin or isochromanquinone. The present sequence is Streptomyces venezuelae sugar (desosamine) biosynthetic gene cluster

Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 U; 0 Other;

Query Match	14.6%	Score 1309.4	DB 6	Length 13613
Best Local Similarity	67.7%	Prod. No. 3,9e-170		
Matches 1850	Conservative	0	Mismatches 876	Indels 5
			Gaps	1
2369	GGGCGGGTTCGACGCGCTTCAGACGCGGGATGACGCGCCCGATGCTGATCCGACCGCTGC			2428
Db	10159	GGGCGGGATTCGGGGCGGGTGGGGATCAGCGGACGAGAACCGCGGGCTCTCCAGCGCTCC		10218
2429	GCGACGTCGCCGTTTCGATTGGTTGAGCCCGGCGGTCACGCACTGCTCGAAACCTGTCAGG			2488
Db	10219	GCGGCGTTCGCGCTTCAGCTGTTTCAGGCGGGCGGTGACACGACTGATCGAAGCGCTCATG		10278
2489	AAGAACTTCGTCGCCCGGCTCGGATGATGCTGTCGGGCGCGACGTCACGAAAGTCGTGAGC			2548
Db	10279	AAGTACTTCGTCGCTCTGACGCGCGGACCTTCGCGCGCTCGACGAAAGTCCTTGAGC		10338
2549	ACCGAATGCGAGGCTCCGGTCCGGGGTGAACCGCGCGGACGATGAGGAGGTGGGCCCGCC			2608
Db	10339	ACCTCGGTGAGGAGGATGTGGGGGTGACGCGGCGCGCATGTGAGCGGATCGCGCGCTCC		10398
2609	AGCCCGGGGAAACCGGCGCTCCCGGTACAGGTAGACGTCCTCCGAGCAGTCGATCTGACCC			2668
Db	10399	AGCTTCGGGAAACCGGCGCTCGCGGTACAGGTACGTCCTCCGCGGAGATCGACTGCACC		10458
2669	GCCACTGTCGGATGAGCGGTTCGGGCGCATGCTTCCGGCGCATCCGCAACAGCTGGGGC			2728
Db	10459	GCGACTGTCGGGATGAGCGGTTCGGGCGCATGCTTCGGGCGGTGATTCGCGAGCAGTTTCGGG		10518
2729	TCCGACACCCGACGCAAGCTCTCCGACAGCGTAAACCGAGGTCCGTCGACGCGCGGGATC			2788
Db	10519	TCCGACCCCGATGCGCAAGCTGTTTCAGGCGCTAACGTAATGCAATGTGAGTCCGGGGATG		10578
2789	CGCTTCGGCGGTGATGACGAAACCGGCGCAGGCGCTTCGCGCAACTCGATCCGTTTCGCC			2848
Db	10579	CGCTTCGGGAAACCGGCTCTCGAAGGGGTGAGGGCGCTTCGAGAGCTCGGCCGCTCTCC		10638
2849	TCCGACAACTGTCGGCTGCTCCGACCGCTGTAGTCTTCGCGGACCGGTGACAAAGTGCAC			2908
Db	10639	TCCGACAGCTTGCCTGCTGACCGGCGCTGTAGTCTTCGCGGATGTTTGAAGAGTGCATC		10698
2909	GCGCGGTGCGGCGCTGACACTGTTGATTCGGCGATGAAAGTCGACGAGTGCATGAGGGCG			2968
Db	10699	GTCCTGCGCTGCGCGGCTGTTGAGGTTCGCGATGAAAGTCGACGAGTGCAGCAGGCGG		10758
2969	TTGGCCGCGCCCGGACGATGATGTGGTTAGGCGCGACCGGACCGGCTCGGCTTGC			3028
Db	10759	GAGGACAGCGCCCGGAGACGATGTAGCGCAAGCCGAGGTGATTCGCGCACTGCGCGCTGC		10818
3029	GCGCGCATTCGACGAAAGTCCCGGAGGTTCTTCGACCCGCTTGAAAGCGGCCGCTTG			3088
Db	10819	GCGGCGAGCTGTGAAAGCGCGGAGGTTCTTCGCGGACCGGCGGAGGCGGCGCTTCTTG		10878
3089	CCGCTGCTGCGCTGATCTGCTGTTTCAAGCCGTCGACGAGGAGGTACGACCGGCGCG			3148



Dp	10879	CCGGTGGATCTGCTCGTACTCTCGTCCGTTTGAAGCCGATAGACGAGGTGGGATGGCGTCC	10938
Oy	3149	AGTTCCTCAACAGCCCGGCTGTGCGCGACGCGTCTGTTCGTGTAGGGCGAAGCGTTGGTG	3208
Dp	10939	AGGCCCCAGAGAGCCCGGCTGTGGCGCTCCAGGGTTCGCTCGGTGTAGCCGAAAGAGTTTCGTG	10998
Oy	3209	TAGACGGGTCAAGGGGAGAGCCCCCGGGGCGCGTGGACACCAACGCCCTCGATACCCGGG	3268
Dp	10999	TAGACGGGTGGCCGCAAGCCGTTGGTCTGTGTGCGTGGCCAGGCTCCCGAGCCGGGG	11058
Oy	3269	TTGGTCAACGGTTTCAAGCCGCGCGAGAGAGGTATCATCGCCGTCGAGTTCCTCCAGCGCACC	3328
Dp	11059	TTGGTGTAGCCGGCTTCAGAGCCCGCCGAGAGATATCATCGCGAGGGAGTTGCCCGGGATATC	11118
Oy	3329	TCGTGATATGACCGAGGTGAGAGATCTCGTTGCCGGGACCAACCGACTCCGCGTGTACTTGG	3388
Dp	11119	TCGTGATATGACCGAGCCGGAACATATGCGCTTTCGGCGGTCAAGGGCGGAGCGGGGTGTAGCGG	11178
Oy	3389	GCACCTTTCACCCCTATACCGAAGATGACACCGGAAATATGACATGCACTCGGCGCGGGGTAGAC	3448
Dp	11179	GCGCCGGGTCAACACGAGACGAGAAGTGTCAACCGAATGTACAGGTCTCGGCGCGGGGTATGAGG	11238
Oy	3449	CCGACGCTGTATACGGGAAAGCGGGGCTCCCGCGGTATACCGCAAGCCCTCAACCGCGCGGGCGG	3508
Dp	11239	CCGACGCTGTATACGGGAAAGAGGGCTTTCGTGGCGAGCGCGCGTGTGAAGACGCGCGCTGT	11298
Oy	3509	TTCAAGCGGCGCTGATCGTGTGTCTCAAGTACTTTCGCGCGGGCGCTGTCTCAACCGCGCTG	3568
Dp	11299	TCGAGCGGGAGAGAGGGGTGTTCTTTCAGATATAGCCCGCGGGGGCGGGTCTTCAGACCGCGGTG	11358
Oy	3569	CCCAAGCTTGGGGATTCGATGTAAACAGGTTCAGACGACGCTCCGCGAACCCGGCGCGGTTCAGG	3628
Dp	11359	CGAGGCTCCGGGACCTTCCCGGAACAGGGCGAGAGAGCGCGCGAAGGCGTCCCGGTTCAGG	11418
Oy	3629	TCGAAACCGAGCGGCATATCTGTCTCAACCGGGGTATACGGCGAGCGCGCGGTATATGTGGCGGTG	3688
Dp	11419	CCCAAGTCTGTAGCGGCGCTCTCTCCAGGGGGGTATAGGGAGCGTGTTCGTATAGGCGACGAGG	11478
Oy	3689	AGTTCGCGCGAGCCGGAACCGCGCTGTCTGTTCGCGCGGTGTCCGATTCGAGATTCGCGCAACCGGTGAGGCGG	3748
Dp	11479	AGCGCGGACGAGATGTATGGCGGGGCGTTCGTTCCGGCTCGTCCGGCGGACACAGCGCCCGCGG	11538
Oy	3749	GTCGACTTCGGCGCTGAGACGCGCCGCAACAGGCGCGGCGCGGTTCGGCGTTCGTGCGCGG	3808
Dp	11539	GCGAGGGGTCTAGCCGGAAGCGGTGGAACGCGCGCCCAATTCGGCTCCGGGATGTGCGGACG	11598
Oy	3809	TCCCGGATCTTCGTCGTGTATACGGCGGGTTCGCTGTAGAGATGTCTTATTCACGTGCGAACC	3868
Dp	11599	CGTTTCGGCGCGGGCGGTGTGGGAAAGGCGCGGGGCGGTCAATCGAGGCGTTCATATCTGTGG	11658
Oy	3869	CTTCTCGCGCTCTGTGTGTGTGCGAGGATCAACGAACCGTTTCCGTTTCTTCACATCCGCG	3928
Dp	11659	GCGTGAATGTCTTGGGGGGGCGCGAGCGGGGCGGGGCGGTGTTCGCGTGGC-----GCCG	11713
Oy	3929	GTTGATCAACGCAACCGCTGTGTGTGTGGCGAATGTGCAATGTGACATCCCGCGGATATTC	3988
Dp	11714	GGTCAATTTCCGCGCGCGGGTTCGCGCAAGACGCAACAGATGTGGGCAACCCGGGGATATTC	11773
Oy	3989	CTCACTGGAACACGAGGGGCGCGGTTCGGAAGGGCGAGCAACCCGTTGGCGAGCGCATCTCGGT	4048
Dp	11774	GTCGTTCGCCATGGCGGGTGTGCGGTTCGGCAGGGACACACAGCGCGCGCGCGCGCTTCGCTCGGT	11833
Oy	4049	CTGTGTCAAGCCGACGCGGATGCGTCCGTTGCGGTGTGAGGCGGTCAATCTGGTGTGCAAGCGGGAGGA	4108
Dp	11834	GTGCGGCAAGCGGGGCGGTTCGAGCTTCCCGCGGTATCGGCTCCAGTCTGTGGCAACCCGGCGCA	11893
Oy	4109	GAACTAGGGCTGTGTGCAACGACTTTCCTCCGCGCGCAAGATTCGCTTGACGCGCGGTTCACGCTC	4168
Dp	11894	GAAATAGGCGCGGGTGTGCAACGCTTTCGCGCTTCAGAGCTTCAGATCAAGAGGTTCGCGGGT	11953
Oy	4169	GATGCCCGGTGTGGCGGTTCCTTCACACAGATGATTCACATCTGTGTATGTTCCTCTCTCTC	4228
Dp	11954	GATGCCCGGTGTGGCGCTTCGTGCATCTTCACAGATCAAGTACTGGTGTGTATTGAGGCGCGG	12013

QY	4229	GGGCGGGAGCGAGTGCACGCGTGAACCGCGCGCTACTCCGCGACGCTCCGCTGTGTAGACCC	4288
Db	12014	GCGGTCGTGTCCGCGACAGAGAGACCGCGGGAGGTCCGACAGGTGTCTCCGCGTAAAGCGCG	1207
QY	4289	GTGGTTGAACCCGGTGTGTGTCTCCCTGTCTCGGTGGGCAGACGCGTCAAGAGGATGAGACCCCAT	4348
Db	12074	GTGGTTGGCGCGGTTCTCCGGTGTGAATGACTCTGGGGAAAACGGGTCAAGGGAGGTAGGCCCAT	1213
QY	4349	GGCGCGGCGCATCTCCGTCATCTTGGCGTTGGTGTCCGATCTCCGTGAACACTTGTCCGG	4408
Db	12134	GGCGCGCGCGCGCTCCGCTCATCTTGGGTGTGTGTCCCGCGCGGGCGTGGCGCGCGGACG	1219
QY	4409	GCCGATGCGCAAGTTGTGTGCATTTGCGCGCGAATCCGTTGGGCGACAGAGCGCGTCTCCGTGTAC	4468
Db	12194	GTCGAAGCGCAAGTTGTGTGAAGGGCGCGAATCCCGGGCGGACAGGTCCGCGTGTGTGTGTAC	1225
QY	4469	CACCGCCCGCGGCTCTGAAAGCGGTGAACCGGCTTGTGTGTGGTGGAGCTGAACAACCTTGGC	4528
Db	12254	GACCGCGCGCGGCTCTGAAAGGGTTTGAAGGCTTGTGTGTGGCTGTGAAGCTGAAGACTTCGGC	1231
QY	4529	GTCAACCGGATCCGCTCCACCGGACGTCCAACCCGTGTGACGCCCAAGGCGTGGCGGCGTTC	4588
Db	12314	GTCCCGGAGGCTGGCGCGCGCGGCGCGCGCGTGCACCGCGCACCGGAGGCGTGTGGCGGCGTCC	1237
QY	4589	GAAAGAAGCGTTTGAACCTGTGTGTGTGGGGGGAGATCTTGGCAGCGCCCTCCACAGTGTCTGG	4648
Db	12374	GAAGTTCACGCGCGCGACGCGCGGTCTGTGTGGGAGACTTTCGACGCTGTGTGGCGCGGCGAGG	1243
QY	4649	TCCGCGCCCAACAGTGTATGCGCGCAAGATCGCGCGCGTCTGTGGGGGTGAACGACCGCGGAC	4708
Db	12434	GCGGCGCCCAAGGTGTGAACGCTCGAAGACGCGCGCAAGGTGTGGGGGTGTGAACCGCGCGCGCAC	1249
QY	4709	GTGGTCCGCGGTGCACCAAGACCGGTCCGCGGCTGCAGCTGCAGAAACACCGGTGTAGTCC	4768
Db	12494	CTGGTCCGCGGTTCGAGTTTGGCGCGGTGTCTCCGGTGCAGTGTCCGCGAAGACCGGGGTGAGGCC	1255
QY	4769	GAGCGACGTCGCGCGCGTGGCGCGGGTGGCGGAGAGTCAATTGAAGGCAATGATCATCTTCACC	4828
Db	12554	GATCCAGCGCATGTGCGGTGGCGGGGTGGCGGAGACGTATCGAGCGCATGTATCATCTTCGCC	1261
QY	4829	GGTGACGTCAACCGGCGCGGACACCAAGTTCAGAGGCGACCGGTGCGTGTGCAGGTGGCGAT	4888
Db	12614	GGTGAAGCGCGCGCGCGCGTGGCGCGGAGAGCTGAAGCCGCGCGTGTGGCTGTGAAGTGTGCAC	1267
QY	4889	GCAGTGCCGTATCCCGAACCAAGGTGTGGGACCCCGGCGCTTGAATCTCCGCAACCAAGGGCCC	4948
Db	12674	GGCATGTCCGAGACCCCGCGGAGCCCGGCGGACGCGCTCTTCGAACTGTGGGAGACGAGGGGCC	1273
QY	4949	GCGGTTGTGTAGCCAGTGTGTGTGTTCAGAGGGCCACTCCAGCGGGGCGACGAAACCGCTGTCCG	5008
Db	12734	GCCGTGTGAACGACCACTGTGTGTGTGAAGGGCGCGGTCAAGCCGCTTCGTAACGCTTGGCGCG	1279
QY	5009	GTCCGCCATGTATCGCGCGCGGCCCAAGTGCAGAGGGGTGCAGAGAACGCTCTCGGACACCGCGAA	5068
Db	12794	GTCAATGCGGTTGGGCGCGGCCCAAGAGAAACGGCTGTGAAAGCGCGGGGCGCGCGAA	1285
QY	5069	GATCGCAAGATCCGCTCGGTACGCGCTTCACG	5099
Db	12854	GAATGCGAGGTCCGATTAAGCGCGCTTTCACG	12884
RESULT 7			
ADL91931			
ID	ADL91931	standard; DNA; 13613 BP.	
XX	ADL91931;		
XX	AC		
XX	DT	20-MAY-2004	(first entry)
DE	Streptomyces venezuelae desosamine gene cluster coding sequence.		
XX	narbonolide polyketide synthase gene; polyhydroxyalkanoate monomer;		

KM medical application; industrial application; ds; gene;  
KM desosamine gene cluster.  
XX  
OS Streptomyces venezuelae.  
XX  
PN US2003194784-A1.  
XX  
PD 16-OCT-2003.  
XX  
PF 15-OCT-2002; 2002US-00271889.  
XX  
PR 17-APR-2001; 2001US-00836821.  
PR 18-MAY-2001; 2001US-00860846.  
PR 18-MAY-2001; 2001US-00861289.  
XX  
PA (SHER/) SHERMAN D H.  
PA (LIUH/) LIU H.  
PA (XUEY/) XUE Y.  
PA (ZHAO/) ZHAO L.  
XX  
PI Sherman DH, Liu H, Xue Y, Zhao L;  
XX  
DR WPI; 2004-119267/12.  
DR P-PSDB; ADL91932.  
XX  
PT New isolated nucleic acid comprising a narbonolide polyketide synthase gene from Streptomyces narbonensis, useful for providing a polyhydroxyalkanoate monomer for medical and industrial applications.  
XX  
PS Disclosure; SEQ ID NO 46; 362pp; English.  
XX  
CC The invention comprises coding sequences for the narbonolide polyketide synthase gene from Streptomyces narbonensis. The DNA sequence of the invention are useful for providing polyhydroxyalkanoate monomer for medical and industrial applications. The present DNA sequence encodes a Streptomyces venezuelae desosamine gene cluster protein which was used in the exemplification of the invention.  
XX  
SQ Sequence 13613 BP; 1858 A; 4732 C; 5093 G; 1930 T; 0 U; 0 Other;  
Query Match 14.6%; Score 1309.4; DB 12; Length 13613;  
Best Local Similarity 67.7%; Pred. No. 3.9e-170;  
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;  
QY 2359 GGGGCGGTGACGCGCTTCAAGCGCGGAGTCAAGCGCCCGATGCTACACCGCTCG 2428  
DB 10159 GGGGGGATCGGGGCGGGTGCAGGTCAAGAGAGAGCGCGGCTCTCCACGCGTCC 10218  
QY 2429 GCGAGCTCCGTTGAGTTGTTGAGCGCGGCGGTCAAGCGATGCTGAGAAACGTCAGG 2488  
DB 10219 GCGGCGTGCCTCCAGCTGTGTTCAAGCGCGGCGGTGAGACCTGATCGAAGCGTCCATG 10278  
QY 2489 AAGAACTGTCCTCCCGCTGCGGATCGATGCTGCGGCGCGAGCTGACGAAGTCTGAGCG 2548  
DB 10279 AAGTACTGTCGCGCTGACGCGCGCACCTCGCGCGCGCTTCACGAAGTCCCTGACG 10338  
QY 2549 ACCGAGTCAAGCTCCGCTGCGGGGTGACCGCGCGGAGAGTACGCGGGTGGCCCCCGCG 2608  
DB 10339 ACCTGATGAGGAGAGTGTGCGGGGTCAAGCGCGCGCGGATGTACGCGGGTGCAGCGTCC 10398  
QY 2609 AGCCCGGGGAAACCGGCTCCCGGATACAGTGAAGTCTCCGAGAGGTGATGTCGACCC 2668  
DB 10399 AGGTGCGGGAAAGCGGCTCCCGGATACAGTGAAGTCTCCGAGAGATTCACCTGACCC 10458  
QY 2669 GCCACTTCGAGTGGGCGGTGCGGCGCATGCTTTCGCGCGGATTCGCAACAGCTGGGCG 2728  
DB 10459 GCGACTGCGGGTGGCGGATGCGGCGCATGCTGCGGGCTTGATCCGAGAGCTTCGCGCG 10518  
QY 2729 TCGAACCCCGACGAGGCTCTCCACGCTTAACCCAGGTGCTGTCGATGCGCGGGGCTC 2788  
DB 10519 TCGGCGCGGATGCGAGGCTGTTCAGGGGTGAGCGGTGATGTGAGTCCGCGGGGTG 10578  
QY 2789 CGCTCGGCGGCGTGTGTCGACGAACGCGGCGAGGCGCTCCCGCACTCGGTCGCGCC 2848

DB 10579 CGCTCGGCGACCCGCTCTTCGAAAGCGTTGAGGAGCTCTCGAGCTCGGCTCTCTCC 10638  
QY 2849 TCGAACAACCTGCGCTGTCGCGACCGCTTAAGTCTCGCGACGCGTGAAGTTCGAGC 2908  
DB 10639 TCGGCAAGCTTGGCTGTGTACAGCGCGCTGTAGTCTCTCGGAAATTTGAGAAATGATC 10698  
QY 2909 GGCCTGTCGCGGCTGAGACTGTTGAGTTTCGCGAATGAAATGACAGAGTGTATGAGCGG 2968  
DB 10699 GTCTGCGCTGCGCGCGGTGTGAGAGTGGCGATGAAATGACAGAGTGTATGAGCGG 10758  
QY 2969 TTGGCCCCCGCGCAGAGATGATGTGTTGAGGTCGAGCGGACCGGTGCGCCCTTCG 3028  
DB 10759 GAGGCAAGCGCCGCGGACAGATGTAGCCGAAGCGGATTGATCGCGACTCGCGCTCG 10818  
QY 3029 GCGCGCATTCGCGAGGAAGTCCCGAGAGTTCTTCCTGACCGGCTCGAAGGCGCACCGCTTG 3088  
DB 10819 GCGCGAGCTGTGTAAGAGCGCGAGTTCTTCGCGACGCGGCGGAAAGCGGCTTCTTG 10878  
QY 3089 CCGGTGATGAGCTGTGTACTGCTGTGTTGAGCGCCGTACAGAGAGTACGGAACCGCGCCG 3148  
DB 10879 CCGGTGATGAGCTGTGTACTGCTGTGTTGAGCGCCGTACAGAGAGTACGGAATGCGGTGC 10938  
QY 3149 AGTCCCAAGGCTGCGGCTGCGCGCGAGCTGTCTGTTGAGGCGGAAAGCGTTGCTG 3208  
DB 10939 AGGCCCCAGAGCGCGGCTGCGCGCTGCTCCAGGATGCGCTGAGACCGGAAAGATTCGTG 10998  
QY 3209 TAGACGATCAAGGCGGAAGCCCGCGCGGCGGCGGACACACAAACGCGCCGATACCGCGG 3268  
DB 10999 TAGAGGATGAGCGCGAGAGCGGTGTGTGTGAGCGGTGCGGAGCTCCGAGGCGCGG 11058  
QY 3269 TTGATCAAGGCTTCCAGAGCGCGCGGAGAGTATACGCGCTCGGATTCCTCGACGCAAC 3328  
DB 11059 TTGATGAGCGGCTTCCAGAGCGCGCGGAGAGTATACGCGCGGATTCCTCGACGCAAC 11118  
QY 3329 TCGTGTATGACCGAGGATGAGATTCGTTGCTGCGGCGGACACACGATCTCGGCTTACTTG 3388  
DB 11119 TCGTGTATGACCGAGCGGAAATGCGGTTCGCGGCGGTGAGGCGGAGCGGATCTGACCGG 11178  
QY 3389 GCACTGTGACCGGCTACGAGAGTGAACAGCGGAAATGACAGTCCGCGCGGAGTACGAC 3448  
DB 11179 GCGCTGAGTACACGAGAGAGAGTGTGACGAGAGTGTGACAGTGTGCGGCGGAGTACGAG 11238  
QY 3449 CCGAGCTGTATACGAGAGCGGCGCTCCCGCGTACCGGACGCTTCAACCGCGCGGCGCG 3508  
DB 11239 CCGAGCTGTATACGAGAGAGAGGCGCTTCTGCGGCGGCGCGGCGGCGGCGCGCTGT 11298  
QY 3509 TTGACGCGCTGATGCTGTGTTGCTTCAAGTACTTGGCCCGCGGCGCTGTCTCAACCGCGGTG 3568  
DB 11299 TCGAGCGGAGCAAGGATGTTCTTCAGATACCGCCCGCGGCGGCGGCTGTCTCAACCGCGGTG 11358  
QY 3569 CCGAGCTGCGGAGAGCGATTCGAAAGAGTTCAGAGCTGCGCGAAACCGGCGCGGCTTCAGG 3628  
DB 11359 CCGAGCTGCGGAGAGCTGCTCCAGACAGGCGGAGAGCGCGCGAAAGAGCTTCCGATTCAGCG 11418  
QY 3629 TCGAACGAGCGGCACTTGTCTTCAAGCGGAGTGAACCGCGAGCGCGGTAGTGGGCGGTG 3688  
DB 11419 CCGAGCTGATGCGGCGCTCTTCAAGCGGAGTGAAGGAGCGGTGCGGTACGACGAGCGG 11478  
QY 3689 AGTTGCGGAGCGCGAGCGCTGTGCTTTCGCGGAGTGTGAGATGCGGACCGGTAGAGCGG 3748  
DB 11479 AGCGGAGAGAGTGGCGGCGGAGTGTTCGCGCTGCTGCGGCGGCGACGAGGCGCGCGCGG 11538  
QY 3749 GTGACCTTCGCGGTGAGAGCGCGCGACACAGCGCGCGGCGGAGTGGGCTGTGCGCGG 3808  
DB 11539 GCGAGGATGCTGAGCGAGCGGCGGTGAGCGCGCGCGCGCGAGATGAGCTCCGAGGATGCGCGAG 11598  
QY 3809 TCCGAGATCTCCGTGAGTACGCGCGGTGCGCGCGGTAGAGTGTCTTCAATGACGTGCGAAC 3868  
DB 11599 CGTTTCGCGCGGCGGCGGTGCGGAAAGCGCGGCGGCGGTCAATCGGAGAGTCCAAATCGTGG 11658  
QY 3869 CTTTCGCGGTGTGTGTGCGAGATCAAGAACGTTGCGTTTCCGCTTGTCCCACTCGCG 3928

Db	11659	GCCTGAGATGTCGTGGGAGGGGGCCGCGAGCGGGGGGGGGGGGGCCGCTGTGGCGGTGGCC-----GCCC	11713
OY	3929	GTTGATCAACGCACCGCTGGTGGTGGGAGAGTCCGAGTACCTCGCACACCCCGCGAGATGTC	3998
Db	11714	GGTCAGTTTCGGCGCGCGCGGTCTCGCGAGAGACCGACGACGATCGGCGCACCCCGCGAGATGTC	11773
OY	3989	CTCACTGGACACCGAAGGGGGCCGGTCCGGAGAGGGCGAGACACCCGTTCCGCGACGCACCTCCGT	4048
Db	11774	GTCCTCCCGATGCGCGGTGGCCGGTCCGGCCAGAGACAGACACGCGCGCGAGGGGTTCCGT	11833
OY	4049	CTGTGTCAAGCCGACAGCGGTGGCTCCGGTCCGGTGAAGGGCGTCACTGTGGTGGCAGGGCGAGGA	4108
Db	11894	GTTGGGAGACGCGGGGGCGTGGCCGGTCCCGCCGTACGGTCCAGCTTCGTGGACGCCGCGCA	11893
OY	4109	GAAATGAGGCTGTGCGACGACCTTCTCCGCGCGCAGGATGCTCTGACGCCGTGACGAGTC	4168
Db	11894	GAAATGAGGCGCGGGTGTGACACGCTTCGGGCTTTCAGGACCTTCATGACAGAGTCCGGTGG	11953
OY	4169	GATCCCGGTGGCGGGTGGCCGTTCCACACAGATGATACGTAATGGTATGTTGCTCTCTCTGTC	4228
Db	11954	GATCCCGGTGGTGGCTCTGTGTGATCTGACGATTCAGTATGTGTGTTGTGAGGCGTGG	12013
OY	4229	GGGCGGAGACGAGATGACACGGGTGACGCGCGCGTACGTCGCGCAGCTCCGCTGGTGAAGAGCGC	4288
Db	12014	GCGGTCGTGTGTGGCGCGACGAGAGACGCGCGGAGAGGTCCGCGAGGTGCTCCGCGTTAGCGCGC	12073
OY	4289	GTTGGTTGACCCGGATTGTGTGCTCCCTGTCTCTCGGCGAAACGCGTCGAGGAGAGTGAAGCCCCAT	4348
Db	12074	GTTGGTTGCGCGGGTTCCTGGGTGATGATCTCGGGAGAAACGGTCTGAGGAGAGTGAAGGCCAT	12133
OY	4349	GGCGCGCGCGCACCTCCGCTCATCTTGGCCGTTGGTGGCCGATCTCGGTACCACTTGTCCGG	4408
Db	12134	GGCGCGCGCGCGCTCCGCTCATCTTGGGCTTGGTGGTCCGCGCGCGCGGCTCCCGCGCGCAG	12193
OY	4409	GCCGATGCCGAAGTTGTGTCATGAGCCCGGATCCGTTTGGGCGACGAGCGCGGTGGTCCGTTGAC	4468
Db	12194	GTCBAAGCTCAAGTTGTGTGAGAGGGCGCGGATTCGCGCGCGCGAGGTCCGCGCTTCGTCCGTGAC	12253
OY	4469	CACCGCGCCCGCTTCGAAGCGGTTGACCCGCTTGTGTGAGCGTGAACCTTGAAACCTCCGCGC	4528
Db	12254	GACGCGCGCGCGCTTCGAAGCGGTTGACCGGCTTGTGTGAGTGAACCTTGAAACCTCCGCGC	12313
OY	4529	GTCACCGGATTCGCGCACCCGACAGTTCACCTCGTGTGACAGCCAGGGCGTGGCGCGCTGC	4588
Db	12314	GTCGCCAAGGCTGGCGCGCGCGCGCTGTGACCGCGCAGCCGAGCGGCGTGGCGCGCGCTGC	12373
OY	4589	GAAAGAGAGGTGACCTGTGTGTGTGGCGCGCGGATCTTCCGACAGCGCTTCACAGAGTCTGG	4648
Db	12374	GAAATGACAGCGCGACCGCGCTGTGTGTGTGAGACCTTTCGCACTGTGTGTGGCGCGCACGAG	12433
OY	4649	TCCGCGCCACAGGTGTATCGCCGACGATTCGCGCGCGGTCTCGCGGGTGTACAGAGCGCGCGAC	4708
Db	12434	GCGGCGCCACAGAGGTGTACCGCCGACGACCGCGCGAGAGTGGGGGTGTACCGCGCGCGCAC	12493
OY	4709	GTTGGTCCGGGTTCGACCAAGACCGGTCGCGCGGTTGACGTGTGCAAAACCGAGTGTAGTTC	4768
Db	12494	CTGTGTCGGGGTTCGAGGTTCGCGGTGTCTCGGGTGAATGTCCGCGAACAACCGGGGTGAGGCC	12553
OY	4769	GAGCCACACTGCGCGCGGTGGCGCGGTGGCGCGGAGGTCACTTGAACGCTATGATCACTTTCAC	4828
Db	12554	GATCCACACGACGTGTCGGTCCGGGTGGCGGGAGAACTGTATTCAGACGAGTATATCACTTCCGC	12613
OY	4829	GTTGACGTTCACCGGCGCGCACGACACCAATTTCAGAGCGAGACGATGTGCACTTGCAGGTGGCAT	4888
Db	12614	GTTAGAGGCGCGGGCGCGTGGCGCGAGAGACTGTGAGACCCCGGCGTGTGGCTTGTACAGTGTGGCAC	12673
OY	4889	GCAGTGCCGTATCCCCCAACAGGTTCCGGCGACCCCGGCGCTTGAACTCCCGCACCAAGGAGCC	4948
Db	12674	GGCATGCGCGAGCCCGGCGAGCCCGGCGACGCGCTCTCGAACTCGCGGACAGAGGGAGCC	12733
OY	4949	GCCGTTGGTGAAGCATGTTGTTTTCAGAGGCCCATCTCAGCGGCGCAGAGAACCGCTGCGC	5008
Db	12734	GCCGTTGACAGGCATGTGCTGTGAGAGGCGCGGTTGACCGCTCGTACAGCTTGAGCGC	12793

OY		5009	GTCGCCCATGCTCGGGCCGCACAGTCAAGGGGTGAAGAAAGCCTCGGACAACGCCGAA	5068
Dd		12794	GTCGATCGCGTTGGGCCCCGCCCAAGAAGACCGCTGATCGAAAAGCGCGGGCCGCCGAA	1285
OY		5069	GATCGCACGATCGGTCGGTAGCGGCTTGACG	5099
Dd		12854	GAATGCAGAGTCGGATTAAAGCGCTTTTACAG	12884
RESULT 8				
AZA87319	ID	AZA87319	standard; DNA; 13613 BP.	
XX	AC			
XX	AAZ87319;			
XX				
DT	15-SEP-2003	(revised)		
DT	05-JUN-2000	(first entry)		
XX				
DE	S. venezuelae desosamine biosynthetic gene cluster pikB.			
XX				
KW	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;			
KW	neomethymycin; narboxmycin; polyhydroxyalkanoate monomer synthase;			
KW	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;			
KW	chronic obstructive pulmonary disease; respiratory inflammation;			
XX	hypercholesterolaemia; crop protection agent; ds.			
XX				
OS	Streptomyces venezuelae; ATCC15439.			
XX				
FH	Key	Location/Qualifiers		
FH	CDS	1..809		
FT		/*tag= a		
FT	/product= "pikB gene cluster protein #1 (AAV77204) "			
FT	/note= "No initiation codon given in the specification"			
FT	CDS	806..2014		
FT		/*tag= b		
FT	/product= "pikB gene cluster protein #2 (AAV80998) "			
FT	CDS	2162..6741		
FT		/*tag= c		
FT	/product= "pikB gene cluster protein #3 (AAV77205) "			
FT	CDS	6834..7402		
FT		/*tag= d		
FT	/product= "pikB gene cluster protein #4 (AAV77206) "			
FT	/transl_except= (pos:6837..6841, aa:Gln)			
FT	CDS	7492..8205		
FT		/*tag= e		
FT	/product= "pikB gene cluster protein #5 (AAV77207) "			
FT	CDS	complement(7942..8205)		
FT		/*tag= f		
FT	/partial			
FT	/product= "pikB gene cluster protein #6 (AAV77208) "			
FT	/note= "No termination codon given in the specification"			
FT	/transl_except= (pos:8270..8272, aa:Val)			
FT	/transl_except= (pos:8273..8275, aa:Thr)			
FT	/transl_except= (pos:8276..8278, aa:Gly)			
FT	CDS	complement(10126..11139)		
FT		/*tag= g		
FT	/product= "pikB gene cluster protein #7 (AAV80999) "			
FT	CDS	complement(11271..12149)		
FT		/*tag= h		
FT	/product= "pikB gene cluster protein #8 (AAV77209) "			
FT	CDS	complement(12342..13799)		
FT		/*tag= i		
FT	/product= "pikB gene cluster protein #9 (AAV77210) "			
FT	/complement(13706..15043)			
FT		/*tag= j		
FT	/product= "pikB gene cluster protein #10 (AAV77211) "			
FT	CDS	15404..15574		
FT		/*tag= k		
FT	/product= "pikB gene cluster protein #11 (AAV77212) "			
XX				
XX	MO200000620-A2.			
NN				

XX 06-JAN-2000.  
PD  
XX 25-JUN-1999; 99MO-US014398.  
PF  
XX 26-JUN-1998; 98US-00105537.  
PR  
XX (MINU ) UNIV MINNESOTA.  
PA  
XX Sherman DH, Liu H, Xue Y, Zhao L,  
PI  
XX WPI; 2000-160679/14.  
DR P-PSD; AAY77204, AAY77205, AAY77206, AAY77207, AAY77208, AAY77209,  
DR AAY77210, AAY77211, AAY77212, AAY80998, AAY80999.  
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.  
PT synthesis of methymycin and pikromycin.  
PS  
XX Diaclosure; Fig 32; 438bp; English.  
XX  
CC The invention relates to an isolated and purified nucleic acid segment  
CC comprising a desosamine biosynthetic gene cluster, a fragment or its  
CC biologically active variant, where the nucleic acid sequence is not  
CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or  
CC streptomycetes antibiotics. The invention also relates to a macrolide  
CC biosynthetic gene cluster, or fragments thereof. The macrolide  
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,  
CC pikromycin, neomethymycin, nadomycin or a combination of these  
CC compounds. Recombinant or augmented cells comprising the desosamine  
CC and/or macrolide biosynthetic gene clusters are useful for the production  
CC of biologically active macrolides. The macrolide biosynthetic proteins  
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and  
CC nadomycin. The alternative termination of polyketide synthesis may be  
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)  
CC monomers. The compounds produced by the recombinant host cells are useful  
CC as biopolymers, e.g., in packaging or biomedical applications, to  
CC engineer PHA monomer syntheses or to prepare biologically active agents,  
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,  
CC chronic obstructive pulmonary disease as well as other diseases involving  
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based  
CC antibiotics which are active against a variety of organisms, e.g.,  
CC bacteria, including multi-drug resistant pneumococci and other  
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop  
CC protection agents (e.g., fungicides or insecticides) via expression of  
CC polyketides in plants. The present sequence represents the desosamine  
CC biosynthetic gene cluster (pik) from *Streptomyces venezuelae* ATCC 15439,  
CC as given in figure 32. (Updated on 15-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 13613 BP; 1858 A; 4733 C; 5092 G; 1930 T; 0 U; 0 Other;  
Query Match 14.6%; Score 1307.8; DB 3; Length 13613;  
Best Local Similarity 67.7%; Pred. No. 6.4e-170;  
Matches 1849; Conservative 0; Mismatches 877; Indels 5; Gaps 1;  
OY 2369 GGGCGGCTGACGCGGTCAGACGGCGGATCAGAGCCCGATGCTGATCCACCGCTCG 2428  
DB 10159 GGGGGGATCGGGGCGGGGTGCGGATCAGCGCAGAGAGCGGGGCTCTCTCCAGCGTCC 10218  
OY 2429 GCGAGCTCCGTTGATGATTGTTAGCGCGGGCGGTACACGATGCTGAGAAACGTTGAGG 2488  
DB 10219 GCGGGGTGCGGCTTCAGCTGGTTCAAGCGGGCGGTGAGCACTGATTCAGAGCGCTGATG 10278  
OY 2489 AAGAACTGTTCCCGCGGCTGCGGATTCGATGCTGCGGCGCGAGCTGACGAAATGCTGAGCG 2548  
DB 10279 AAGTACTGTCGCGCTTCAGACGCGCACCTGCGCGCGCGCTTCACAGCAAGTCCCTGACG 10338  
OY 2549 ACCGAGTGGAGGCTCGGATCGGGGATGACCGGCGCGGAGCGATGACGGGATGAGCGCGCGC 2608  
DB 10339 ACCTGCGTGAAGGAGATGTCGGGGATCACGCGGCGCGGATTAACGGGATCGCGCGCTCC 10338  
OY 2609 AGCCCGGAGAAACCGGCGCTCCCGGTACAGGTAGACGCTCTCCGAGCAGGTGATCTGCAAC 2668  
DB 10399 AGGTGCGGAGAGCGCGGCTCCGGGTACAGGTACAGTCCGAGAGAGATTCGACCTGCAAC 10458

OY 2669 GCCACCTGGAGATGGGCGGATCGGGCGCATGCTTTCGCGCCGATTCGCAACAGCTGGCGC 2728  
DB 10459 GCCACCTGGCGGATGGCGGATCGGGCGCATGCTTTCGCGCCGATTCGCAACAGCTGGCGC 10518  
OY 2729 TCGACACCGCCGACGCGAGGCTCTCCAAAGCGGTAAACCAAGTGGGTCTGCAATGCCGGGATC 2788  
DB 10519 TCGGCGCGGATGGCGAGCTGTTTCAAGGCGCTAACCGGTAGTGTGAGTCTCCGGGGGTG 10578  
OY 2789 CGCTCGCGCGGATGATGATGCAAGAAACCGGGGAGGCGCCCGCGCACTCGGTCCGTGCGCC 2848  
DB 10579 CGCTCGGAGACCGGCTCTCTGAAAGGCTTGAAGGAGCTCTGAGAGCTCGAGCCCGCTCTCC 10638  
OY 2849 TCGGACAACTGCGCTGCTCTCCGACGCGTGTAGTCTCTCGCGGACGCTGACGAAGTCGAGC 2908  
DB 10639 TCGGCGAGCTTGCCTCTGTATACGCGCGCTGTATCTCTCGGAAATGTTAGAGATGATC 10698  
OY 2909 GGGCGGTGGGAGCTGCACTGTTGAGTTGCGGCGATGAAAGTTCACACAGGTGATGAGGCGG 2968  
DB 10699 GTCCTGCGCTGCGCGGCTGTTGAGGTGCGCGATGAAAGTTCACACAGGTGAGAGGCGG 10758  
OY 2969 TTGGCGCGCGCGCGGACGAGATGATGTTGAGGCGGAGCGGACCGGTCGCGCGCTTCG 3028  
DB 10759 GAGGCAAGGCGCGGAGACAGATGTAGGCGAAGCGAGTTGATCGGCGACTGCGCTTCG 10818  
OY 3029 GCGGCGATTCGCGAGAAAGTCCCGGAGGTTCTTCTCTGACCCGTTTCGAAAGCGCGCACGCTTG 3088  
DB 10819 GCGCGAGCTGCTGAGAAACCGGCGCAAGTTCTTCGCGACGCGGAGAAAGGCGGCTTCTTG 10878  
OY 3089 CCGGTGATGAGCTGATGATCTGCTGATGTTTACGCCCGTACAGGAGATGACGACCGCGCGC 3148  
DB 10879 CCGGTGATGATGCTGATGATCTGCTGATGATGATGATGATGATGATGATGATGATGATG 10938  
OY 3149 AGGTCCCAACGCGCGGCTGCG 3208  
DB 10939 AGGCGCCAGAGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10998  
OY 3209 TAGACGCTGACGAGGAGGAGGCG 3268  
DB 10999 TAGACGCTGAGCG 11058  
OY 3269 TTGATGACGAGGCTGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3328  
DB 11059 TTGATGACGAGGCTGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11118  
OY 3329 TCGTGTATGACCGGAGTGAAGATCTCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3388  
DB 11119 TCGTGTATGACCGGAGTGAAGATCTCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11178  
OY 3389 GCACCTGTACCGGTAAGGAGAGTGAACGAGGAGATGACGAGGAGATGACGAGGAGATGACG 3448  
DB 11179 GCACCTGTACCGGTAAGGAGAGTGAACGAGGAGATGACGAGGAGATGACGAGGAGATGACG 11238  
OY 3449 CCGAGCGTGTACCGGTAAGGAGAGTGAACGAGGAGATGACGAGGAGATGACGAGGAGATGACG 3508  
DB 11239 CCGAGCGTGTACCGGTAAGGAGAGTGAACGAGGAGATGACGAGGAGATGACGAGGAGATGACG 11298  
OY 3509 TTGAGCGGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3568  
DB 11299 TTGAGCGGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11358  
OY 3569 CCGAGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3628  
DB 11359 CCGAGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11418  
OY 3629 TCGAACCGAGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3688  
DB 11419 CCGAGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11478  
OY 3689 AGTTGCGGAGAGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3748  
DB 11479 AGCGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11538

[illegible][illegible]



```
FT CDS 47794..49083
FT /*tag= k
FT /product= "Polyketide synthase #11"
FT 49092..49814
FT /*tag= 1
FT /product= "Polyketide synthase #12"
FT 49868..51226
FT /*tag= m
FT /product= "Polyketide synthase #13"
FT /transl_except= (pos:51221..51223, aa:Xaa)
FT /note= "Xaa corresponds to amino acids from position 452-
FT 1811"
FT CDS 51506..53416
FT /*tag= n
FT /product= "Polyketide synthase #14"
FT /note= "CDS does not include start codon"
FT /partial
FT CDS complement(54569..53358)
FT /*tag= o
FT /product= "Polyketide synthase #15"
FT 54897..56342
FT /*tag= p
FT /product= "Polyketide synthase #16"
FT 56408..57634
FT /*tag= q
FT /product= "Polyketide synthase #17"
FT 57657..59123
FT /*tag= r
FT /product= "Polyketide synthase #18"
FT /note= "CDS does not include start codon"
FT /partial
FT CDS 59163..60196
FT /*tag= s
FT /product= "Polyketide synthase #19"

CA2391131-A1.
PN 19-NOV-2002.
PD 26-JUL-2002; 2002CA-02391131.
PF 26-JUL-2001; 2001US-0307629P.
PR 26-JUL-2001; 2001US-0307629P.
PX (ECOP-) ECOPIA BIOSCIENCES INC.
PY Yang X, Staffa A, Parnet CM;
PA
PE WPI; 2003-343556/33.
PF P-PSDB; AAE36995, AAE36996, AAE36997, AAE36998, AAE36999, AAE37000,
PR AAE37001, AAE37002, AAE37003, AAE37004, AAE37005, AAE37006, AAE37007,
PX AAE37008, AAE37009, AAE37010, AAE37011, AAE37012, AAE37013.
PY
PT Novel isolated polypeptide involved in biosynthesis of macroides by
PT microorganisms, useful for biosynthesis of macroides by microorganisms,
PT preferably for biosynthesis of rosamycin.
PX
PS Example 2; Page 59-94; 206pp; English.
XX
XX The invention relates to genes and proteins involved in the biosynthesis
XX of macroides by microorganisms. In particular it relates to the nucleic
XX acids forming the biosynthetic locus for rosamycin (a 16-member
XX macroide antibiotic) from Micromonospora carbonacea. The invention is
XX useful for the biosynthesis of macroides and related chemical structures by
XX direct manipulation of macroides and related chemical structures by
XX chemical engineering of the proteins involved in the biosynthesis of
XX rosamycin. It is useful to catalyse certain biochemical reactions, in
XX vitro or in vivo, to direct or enhance the synthesis or modification of a
XX polypeptide, polypeptide substrate or its precursor. The present sequence
XX is M. carbonacea polypeptide synthase. (PKS) type I gene cluster. (Updated
XX on 27-OCT-2003 to standardise OS field)
SQ Sequence 60196 BP; 6934 A; 20042 C; 23822 G; 9398 T; 0 U; 0 Other;
```

```
Query Match 13.9%; Score 1241.2; DB 10; Length 60196;
Best Local Similarity 63.2%; Pred. No. 6,7e-161;
Matches 1979; Conservative 0; Mismatches 1138; Indels 15; Gaps 4;

QY 2395 GCGGATCAGGCCCCCGATGCTGATCCCAACCGGTGGGAGCTCCCGTTGAGTTGAG 2454
DB 59120 GCGGAGGAAACCCCGTCCGCTCCCAAGCATGGGAGATCGATCTCCATCTGTTGAG 59061
QY 2455 CCGGGGCGTACCGGATCGGTGAAACCGGTGAGGAAAGAACTGTCCTCCCGTGGGATC 2514
DB 59060 CCGGCGGATACCGGCTGATCGATTCGATTCAGGAAATTTCTGTCAGGCTTCGGGAC 59001
QY 2515 GATCTGCGGCGCCGACGTGACGAAGTCTGGAAGACCGAGTGAAGCTTCGGTCCG 2574
DB 59000 CACGATCCGCGCGGTGTGACGAACGTCACACACCTGCGGCAACGTCGTGCGGAGGA 58941
QY 2575 GACCCGCGCGGACGATGAGCGGTGCGCCCGCCGACGCGCGGAAACCGGCTTCGCGTA 2634
DB 58940 TACCGTCCGATCGATGAGCGGTGCGGACCGGCGGAAACCGGCTTCGCGTA 58881
QY 2635 CAGGTAGACGTCCTCCGACGAGTCTGATCTGACCGGACCGGATGGCGGTGCGGCG 2694
DB 58880 GAGGTAGACATCAACCGAGATATCACCTGACCGACACCTGCGGAGTGGCGCGAGGCG 58821
QY 2695 CATCGTTTCGCGCGGATTCGCAACAGCTGAGGCTGACACCCCGACGAGGCTTCGCA 2754
DB 58820 CATGCTCCGCGCGGATGACGACGAGCTCAAGTCCCTCCGTCATCAGGCTGTGCAAG 58761
QY 2755 CCGGTAAACCAAGTCTGCTGATGCGCGGAGTTCGCTCGGCGCGGTATGTCAGAAACG 2814
DB 58760 GCGGTACCGGTAGTCATGTCAGGAGTTCGCTGATGTCCTTCCTCCGAAACCG 58701
QY 2815 GCGAGGCGCTCCCGCAACTCGATCGGTGCGCTCGGACCAACTGCGTCTGTCGAC 2874
DB 58700 GTGCACTTCGAGCTGAGAGCTCGGCGGCTGTCCAAGGAGAGCTTCGCTCGGCGCGCC 58641
QY 2875 GCTGTAGTCTCTCGGCGAGTGTGAGGAGTGAAGCGGCGGCTGCGGCTGCACTGTTGAG 2934
DB 58640 GCTGTAGTCTCTCGGCGAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 58581
QY 2935 TTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2994
DB 58580 CTGCGGACGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 58521
QY 2995 GTTGAAGCGGACCGGACCGGATGCGGCGGCTTCGCGGCGGATTCGCGAGAAAGTCCG 3054
DB 58520 GCTGAAGCGGACCGGACCGGATGCGGCGGCTTCGCGGCGGATTCGCGAGAAAGTCCG 58461
QY 3055 GTTCTTCCTGACCGGATGAGGCGGACCGGATGCGGCGGATTCGCGAGAAAGTCCG 3114
DB 58460 GTTCCGCGGATGAGGCGGAGGCGGCGGATGCGGCGGATTCGCGAGAAAGTCCG 58401
QY 3115 GTTCAAGCGGATGAGGAGGATGAGGAGGCGGCGGATTCGCGAGAAAGTCCGCGGCG 3174
DB 58400 GTTCAACCGGATGAGGAGGATGAGGAGGCGGCGGATTCGCGAGAAAGTCCGCGG 58341
QY 3175 CAGGCTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3234
DB 58340 GAGCTTCTGCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 58281
QY 3235 GCGCGGCTGACACCAACGCGCGGATGCGGCGGATTCGCGAGGCGGCGGAGGCGGAG 3294
DB 58280 GCGCGCGGCTGACACGAGTGAACCGAGGCGGCGGATTCGCGAGGCGGCGGAGGCGGAG 58221
QY 3295 GAGGTACATGCGCGGCTGCGGCTTCGCGAGGCGGAGGCGGATTCGCGAGGCGGAGGAG 3354
DB 58220 GACGTACACCGCGGCTGCGGCTTCGCGAGGCGGAGGCGGATTCGCGAGGCGGAGGAGGAG 58161
QY 3355 GTTCCCGGCGGACCGGATTCGCGGCTTCGCGAGGCGGAGGCGGATTCGCGAGGAGGAG 3414
DB 58160 GTTCCCGGCTGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGAGG 58101
QY 3415 ACAGCGGAAATGACGTCGCGGCGGAGGAGGCGGAGGCGGAGGCGGAGGCGGAGGAGG 3474
```





ID ADM45913 standard; DNA; 84428 BP.  
XX  
AC ADM45913;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Streptomyces mycarofaciens midcamycin polyketide synthetase DNA.  
XX  
KM macrolide; midcamycin biosynthesis; polyketide synthetase; PKS;  
XX antimicrobial; ds; gene.  
OS Streptomyces mycarofaciens.  
XX  
PN JP2004049100-A.  
XX  
PD 19-FEB-2004.  
XX  
PF 19-JUL-2002; 2002JP-00210516.  
XX  
PR 19-JUL-2002; 2002JP-00210516.  
XX  
PA (MEIJ ) MEIJI SEIKA KAISHA LTD.  
XX  
DR WPI; 2004-174115/17.  
XX  
DR P-PSDB; ADM45914, ADM45915, ADM45916, ADM45917, ADM45918, ADM45919,  
XX ADM45920, ADM45921, ADM45922, ADM45923, ADM45924, ADM45925, ADM45926,  
XX ADM45927, ADM45928, ADM45929, ADM45930, ADM45931, ADM45932, ADM45933,  
XX ADM45934, ADM45935, ADM45936, ADM45937, ADM45938, ADM45939, ADM45940,  
XX ADM45941, ADM45942, ADM45943, ADM45944, ADM45945, ADM45946, ADM45947,  
XX ADM45948, ADM45949, ADM45950, ADM45951, ADM45952, ADM45953, ADM45954,  
XX ADM45955.  
XX  
PT New polynucleotides encoding proteins important in the biosynthesis of  
PT midcamycin.  
XX  
PS Claim 2; SEQ ID NO 1; 264bp; Japanese.  
XX  
CC The invention relates to a novel polynucleotide, specifically a  
CC biosynthesis gene, comprising a nucleotide sequence which encodes a  
CC protein, where the gene is concerned with the biosynthesis of  
CC midcamycin. The polynucleotide of the invention demonstrates a  
CC antimicrobial activity and may be useful for the biosynthesis of  
CC midcamycin, as well as for manufacturing macrolide compounds other than  
CC midcamycin. The current sequence is that of the Streptomyces  
CC mycarofaciens midcamycin polyketide synthetase (PKS) DNA of the  
CC invention.  
XX  
SQ Sequence 84428 BP; 11456 A; 29098 C; 30275 G; 13599 T; 0 U; 0 Other;  
Query Match 13.5%; Score 1203.2; DB 12; Length 84428;  
Best Local Similarity 65.5%; Pred. No. 1e-155;  
Matches 1778; Conservative 0; Mismatches 933; Indels 5; Gaps 1;  
QY 2387 CAGACGGCGCGATGATGAGCCCGGATGATCCACCCGCGCGGACGTCCTCGATG 2446  
DB |||||  
QY 3232 CAGGCTATCAAGAAAGACCGCGTGTCCGCCACCGCTCGGCGGATTCACCTTCAGC 3291  
DB |||||  
QY 2447 TGGTTGAGCCCGGCGGTCAACGACTGTGCGAAACCGTGAAGAAACTCGTCCCGCGC 2506  
DB |||||  
QY 3292 TGGTTCAACCGTGTGATGACGACCTGTGCAACCGTCAATGAATACCTGTGCGCTCG 3351  
DB |||||  
QY 2507 TGGGATGATGCTGCGGCGCGGACGTCGACGAACTGTGACGACCGAGTGCAGCTCCG 2566  
DB |||||  
QY 3352 GCCGAGGAATCTGCTCCACCGCTGTCAAAAGCGCTCAACACCTGGGTACGCTTTT 3411  
DB |||||  
QY 2567 TCCGGGTGACCCCGCGCGACGTAAGCGGTGACCCCGCGGCGGAGAAACCGGCGC 2626  
DB |||||  
QY 3412 CCGGCGGAGACCTTGCCGATGTGTAGCGGTCCGCTCCGGGACGTCGCGGAAACCGGCGC 3471  
DB |||||  
QY 2627 TCCGCTACAGGTACGCTCTTCCAGCAGTGTGATCTGACCGCACTCTGCGGATGCGCG 2686  
DB |||||  
QY 3472 TCGGTAGAGGTAGACGTGCGCCAGCAGATGACCTGGAATGCACTCGCGGATGCGCG 3531  
DB |||||

QY 2687 GTCGGGCGATCGTTTCCGCGCGAATCCGCAACGCTGGCGTGCACCCCGACGCAAG 2746  
DB |||||  
QY 3532 GTCGGCGCATGTGTTCCGCGCGAATCCGCAACGATTCGCGCTCCGCGCGGAGAG 3591  
DB |||||  
QY 2747 CTCTCCAAACCGGTACCCCAAGTGGTCTGTGATCCCGGGGTCCGCTCGGCGGTATGCG 2806  
DB |||||  
QY 3592 CTGTTACGCGGTACCCGTAAGTGCACGTGAGGTGGTGTCTTGGCCGGAACCTTCTCC 3651  
DB |||||  
QY 2807 ACGAACCGGCGGAGGCGCTCCGCAACTCGATCGTTCGCTCGACCAACTGCGCTCG 2866  
DB |||||  
QY 3652 CGAAAGCGGTACGCGCTTCTGTAGAGTCCGCGCGCTCGGACCGGACGAGAGCGGTGCG 3711  
DB |||||  
QY 2867 TCCGACCGGTATGATCTTCGCGGACGTTACGAAATGACGCGCGGTGCGGCTGAC 2926  
DB |||||  
QY 3712 GCGCGCGCGGTATGATCTCAACGAGGTACGAAATGATGATGATGATGATGATGATGATGAT 3771  
DB |||||  
QY 2927 TCGTTGAGTTCCGCGAATGAAAGTGCACGAGTGCATGAGGCGGTTGGCGCGCGCGGAG 2986  
DB |||||  
QY 3772 GCGTTCACTCGGAGTAAGTGCACGAGGTGCAGAGCGCGCGCGCGCGCGCGCGCGCG 3831  
DB |||||  
QY 2987 ATGATGTGTTGAGGCGGACCGGACCGGATGCGCGCGCTTCGCGCGCATTCGCGAGAG 3046  
DB |||||  
QY 3832 ACGATGATGTTCAAGCGCGGATTCACCGGTTCCGCGGTGCGGAGCGCGCGCGCGCG 3891  
DB |||||  
QY 3047 TCCGAGAGTTCTTCTGACCCGTTGCAAGGCGCACGCTTCGCGGTGCGCTGATC 3106  
DB |||||  
QY 3892 CGCTCCAGCTGCGGCTTCAACCGCTCCGAAAGCGCGCTTCTTCTGTTGTTGCGCGAT 3951  
DB |||||  
QY 3107 TCGTCTGTTTCAAGCGCGGTACAGGAGGTACGAGCGCGCGGAGGTCCACAGCGCGGAG 3166  
DB |||||  
QY 3952 TCGCATGTTCACCGGTACAGGAGGTTCGACCGCGGTGACGCGCGCGCGCGCGCGCG 4011  
DB |||||  
QY 3167 TGGCGGCGCAGCGTCTGTTGAGGCGGAGCGTTGTTGATGACGTTGAGGCGAG 3226  
DB |||||  
QY 4012 TGCCTTTCACGCTCCGCTCGGTACGAGCGGAGCGCTTGTGTACAAAGTATGAGCGGAA 4071  
DB |||||  
QY 3227 CCGCGCGCGGCGGCGGTGACACCAACGCGCGGATCCCGGTTGTACCGTTCCAG 3286  
DB |||||  
QY 4072 CCGCGCGCGGCGGCGCTCCGAAAGAGGAGCGCGCGGATGTTGAGCGCGCTCCGAG 4131  
DB |||||  
QY 3287 CCGCGGAGAGTACATCGCGGTGCTTCGCAACGCGCTGTTGATGACGAGAGT 3346  
DB |||||  
QY 4132 CCGCGGACAGTACATGCGGTGCGGAGTGTCCGTTGACATGTCATGAGCGCGCG 4191  
DB |||||  
QY 3347 AGATCTGTTGCGGCGACACCGACTCCGCTGATCTTGGACCTGTGACCGGTACG 3406  
DB |||||  
QY 4192 AACATCTTGTGCGGTGCTGACGAGGCGGAGTGTGATGAGCGGCGCGGTACCGGTACG 4251  
DB |||||  
QY 3407 CAGAAATGACAGCGGAAATGACGATGCGGCGCGGATGAGCGCGGATGAGCGGAG 3466  
DB |||||  
QY 4252 CAGAAATGACAGCGGAAATGACGATGCGGCGCGGATGAGCGCGGATGAGCGGAG 4311  
DB |||||  
QY 3467 GCGGCGTCCGCGGTACCGGACCTCCAAACGCGCGGCGGCTTCAAGCGCGGTATG 3526  
DB |||||  
QY 4312 ACCGCTTGTGTGCAAAAGCGGTGCAAAAGCGCTTTCGTTCCAGCGGAGAGCGGTG 4371  
DB |||||  
QY 3527 TTGCTCAAGTACTTGCAGCGCGCGCTGCTCAACCGCGCTGCGGAGTGTGCA 3586  
DB |||||  
QY 4372 TTGCTCAAGTACTTGCAGCGCGCGCTGCTCAACCGCGCTGCGGAGTGTGCA 4431  
DB |||||  
QY 3587 TCGAACAGGTGACAGCTGCGCAACCGCGCGGTTCAGGTGAAACGAGCGCATC 3646  
DB |||||  
QY 4432 CCGAAAGGTGACAGAGCGCGGAGCAATGATGATGATGATGATGATGATGATGATGAT 4491  
DB |||||  
QY 3647 TGTCTCAACGAGGTGAAAGCGGAGCGCGGTGATGAGGTGAGTTCGCGCGGAGCG 3706  
DB |||||  
QY 4492 TTTTTCAGCGGAGTGAAGGCGCGCGCGGTGATGATGATGATGATGATGATGATGAT 4551  
DB |||||  
QY 3707 GCGGTGCTTTCGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3766  
DB |||||  
QY 4552 GCGGCTTTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4611  
DB |||||  
QY 3767 GCCGCGACACAGCGCGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGG 3826  
DB |||||

```

Db      4612 GCCTCGAAGGCGCGCGGATCACTCCGGGAAGGGTGCAGACCGCTACGGTGTGCC 4671
Qy      3827 ACCGCGGTGCGCGGAGAGTGTCTTATGACGTGTGGAACCCCTTCTGGCGTCTGTGTG 3886
Db      4672 ACCGCTTCTTGCCTCACCATGATGATCCACCCCTTTCATTCGAAATGTCTTGTGATG 4731
Qy      3887 CGAGATCA-----CGAACGTTTCCGTTTCCGTTTCCCACTCCGCGTTGATCAAGCA 3941
Db      4732 AGAAGGCGCGTCCGCTGTGTACTGTAGTGGCCCTGAGCCGACCGGCGGTCTACGTCCGCC 4791
Qy      3942 CCGTGTGTGTGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4001
Db      4792 CCGGCGGTGTGCGCGCAACCGGACGATGTCCGATCCCGCGGATATCTCATGTGGAAGCC 4851
Qy      4002 GAGGCGCGGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4061
Db      4852 GTGGAACCGGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4911
Qy      4062 AGCGGTGTGCTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4121
Db      4912 GCGTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4971
Qy      4122 GCGACGACCTTCTTCCGCGCGCAAGATGCGCTGCACCGGTCACCGGTGATGCGGTGAGC 4181
Db      4972 GCCACGACGTTTCTCGGCGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5031
Qy      4182 GTGCCGTCACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4241
Db      5032 TCCCTGTGATCTGACAGACGACGATGATGATGATGATGATGATGATGATGATGATGATG 5091
Qy      4242 TGCACGTCAGCGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4301
Db      5092 AAGACGCGCACACCGGTACTCCGAGAGATTCGTCGAGTGTCTCATGTGTCTTTTG 5151
Qy      4302 TTGTGCTCCCTGTCTTCCGGAACGCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4361
Db      5152 TTGTGCGTCGACGACTTCTCATAACAGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5211
Qy      4362 TCGCTCATCTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4421
Db      5212 TCGCTCATCTTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5271
Qy      4422 TTGTGTCATGCGCGGATCGTGTGCGGACGCGCGTCGTCGTCGTCGTCGTCGTCGTCG 4481
Db      5272 TTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5331
Qy      4482 TCGAAGCGGTGACCGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 4541
Db      5332 TCGAAGGAAATGACGACTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5391
Qy      4542 CCCACCGGACGTCACCGCTGTGTCAGCCGAGGCGGTGCGGCGGTGGAAGAGCTTG 4601
Db      5392 CCACAGAGTCTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5451
Qy      4602 ACCTGTGTGTGCGGCGGATCTTCCGACGCGCTTCACAGTGTGTGTGTGTGTGTGTGTG 4661
Db      5452 ATCCCGGCTTCCCGGCGACGCTCTCCAGGCGGTGTCACATGACGCGCTTACCCACAGA 5511
Qy      4662 TGTACGCGGACGATCGCGCGGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4721
Db      5512 TGCACCCCGAATATCCGAGAGTGTGCGGCGGTGATGCGGCGCGGACCGGCTGGGATCC 5571
Qy      4722 ACCAAGACCGGTGCGCGGTGTGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4781
Db      5572 AACAATTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5631
Qy      4782 GCGTGTGCGGTGTGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4841
Db      5632 GCGTGTGCGGTGTGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 5691
Qy      4842 GCCCGACGACGATTCAGAGGCGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 4901

```

```

Db      5692 GCGTGTACAGTAGTGTGAGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 5751
Qy      4902 CCGACAGGTGTGCGGACCGCGGCTCGAATCTCCGACACGAGGCGCGCGGTCGTGTGAGC 4961
Db      5752 CCGGCGAGTGTGCGGACCGCTTCTGAACTCTCGGTCGAGGTCGCGGTCGTGTGAGC 5811
Qy      4962 CAGTGTGTGTGAGGCGCCATCCGACGCGGACGAGAACGTCGTCGTCGTCGTCGTCGTCG 5021
Db      5812 CACTGTGTGTGACGCGCCGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5871
Qy      5022 GCGCGGCGCAGTCGACGAGGTCGAGAACGCTTCCGACCGCGGAGATGCGCAGATCG 5081
Db      5872 GGGCGGCGCACAGAGCGGCTGAGAGGATGTTGGCGCGCCCGCGAACGCGGAGATCG 5931
Qy      5082 GTGCTATCGGCTTCA 5097
Db      5932 CCGAGTTCGCTTCA 5947

RESULT 11
AAA75638/c
ID AAA75638 standard, DNA, 1565 BP.
XX
AC AAA75638;
XX
DT 22-JUN-2001 (first entry)
XX
DE Nucleotide sequence of the ORF of the pickV gene.
XX
KM Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KM antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KM picromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PE 27-MAY-1999; 99US-00320878.
XX
PF 30-APR-1997; 97US-00846247.
PR 06-MAY-1998; 98US-00073538.
PR 28-MAY-1998; 98US-0087080P.
PR 28-AUG-1998; 98US-00141908.
PR 22-SEP-1998; 98US-0100880P.
PR 08-FEB-1999; 99US-0119139P.
PR 20-MAY-1999; 99US-0134990P.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value.
XX
PS disclosure; Col 45-48; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compound of
CC the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode a
CC C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity of
CC a compound relative to the unhydroxylated compound. The recombinant host

```

CC cells are useful as genetic systems that allow rapid engineering of the  
CC narbonolide polyketide synthase. These would be valuable for creating  
CC novel ketolide analogs for pharmaceutical applications

XX Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;

Query Match 8.2%; Score 729.4; DB 3; Length 1565;  
Best Local Similarity 68.1%; Pred. No. 7.5e-91;  
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

```

Qy 2369 GGGCGGCTCGACCGCTTACAGACCGCGGATCAAGCCCGGATCTCATCCACCGCTG 2428
Db 1530 GGGGGGATCGGGGCGGGTGCAGCGTCAAGCGCGGGCTCTCCACGCGCTCC 1471
Qy 2429 GCGACGCTCCGCTTCAAGTGGTTGAGCGCGGGCTCAACGATCTGTGAAACGCTGAGG 2488
Db 1470 GCGCGCTGCGCTCAAGTGGTTGAGCGGGGCTGAGCGAGCTGATGAAAGCGCTCAAG 1411
Qy 2489 AAGAACTGCTCCCGCGCTGCGATCGATGCTGCGCGCGCGACGTAAGTCTGTGACG 2548
Db 1410 AAGTAAGTCTGCTGCGTGAACGCGCGCGCGCGCGCTGCGAGAGTCCCTGACG 1351
Qy 2549 ACCGATGTCAGAGCTCCGCTCGGGGTGACCGCGCGCGAGCTGACGGGTGCGCGCG 2608
Db 1350 ACTCGGTGAGGAGGTGTGCGGGGTGACGCGCGCGCGAGTGAAGCGGTGCGCGCTCC 1291
Qy 2609 AGCCCGGGGAAACGGGCTCCCGGTACAGGTAGAGTCTCCGAGACGGTGTGACG 2668
Db 1290 AGGTGGGGAAGCGGCTCGCGGTACAGGTACAGTGCAGAGAGATGCACTTGACG 1231
Qy 2669 GCCACTGCGGATGGGCGGTGCGGCGCATCGTTTCGCGCGGATCCGCAACGCTGAGCG 2728
Db 1230 GCGACTGCGGGTGGGGTGGGCGCATGCTGGGGCTTGATCCGAGAGTTCCGCG 1171
Qy 2729 TCGAACCCCGACGAGGCTCTTCAACCGGTACCCAGGTGCTGTGATCCCGGGTCC 2788
Db 1170 TCGGCGCGGTGCGCAGGCTGTTCAGGCGGTAGCGGTGATGTGAGTCCGGGGTGG 1111
Qy 2789 CGCTCGGCGGTGATGTCAGACCGGCGAGGCGCTCCCGCACTCGGCTCGTTCCGCC 2848
Db 1110 CGCTCGGCGACCGGCTCTCTCAAGGCGGTGAGGCGCTCTGAGAGTCCGCGCGCTCC 1051
Qy 2849 TCGGACAACTGCGCTCGTCCCGACCGGCTGATCTTCGCGCGAGCTGACAGTCCAGC 2908
Db 1050 TCGGCGAGCTTCCGCTGTCACGCGCGGTATGCTTCGCGAATGTTGACGAAGTCATC 991
Qy 2909 GCGCGGTGCGGCTGAGTCTGTTGATGTCGCGGATGAAGTGCACCGGTGATGAGCGCG 2968
Db 990 GTCCCTGCGCTCGCGCGTGTGAGTGCAGTGAAGTGCACCGGTGATGAGCGCG 931
Qy 2969 TTGGCCCGGCGCGGAGATGATGCTGTTGAGCGCGACCGGACCGGCTCGCGCTTCG 3028
Db 930 GAGGCGACCGCGCGGAGACGATGATGAGCGGAGGCTTATGCGGACCTCGGCTCG 871
Qy 3029 GCGGCGATCCGCGAGGATGCCCGAGGTTCTTCTGACCCGTTCAAGGCGCGACGCTTG 3088
Db 870 GCGGCGAGCTGCTGAAACCGCGCGAGTTCGCGGAGCGCGCGGAAAGCGCGCTTCTTG 811
Qy 3089 CCGGTGCTGCGCTGATCTCGTCTGTTCAAGCGGCTGACAGGAGTACGACCGCGCG 3148
Db 810 CCGGTGCTGCTGATCTCGTCTGTTGAGCGGCTGAGAGGAGTGCAGGCTGCGTGC 751
Qy 3149 AGGTCCGACAGCGCGGCTGCGCGGCGCGGCTGTTGAGGAGGCGGAGGCGGTTGCTG 3208
Db 750 AGGCGCCAGAGCGCGGCTGCGGCTGCGAGGCTGCGTGTAGGCGGAGAGGATGCTG 691
Qy 3209 TAGACGCTCAGGCGGAGCGCGCGCGGCGGCTGCGACGAAGCGCGCGATCCCGG 3268
Db 690 TAGACGCTGCGGCGGAGCGCGGCTGCGGCTGCGGCGGCGGCTCCGAGGCGGCGG 631
Qy 3269 TTGCTCAACGCTTCCAGGCGCGCGGAGAGTACATGCGCTGCGGCTTCTCGACGCGCAC 3328
Db 630 TTGCTCAACGCTTCCAGGCGCGCGGAGAGTACATGCGCGGAGGCTTCCCGGCTGTC 571

```

```

Qy 3329 TCGTCATGACCGAGGTGAGGATCTGTTGCGCGGACCAACGACCTCCGCTGACTTG 3388
Db 570 TCGTCATGACCGAGCGGAACATGCGTTCCGCGCTGAGGCGGACGCGGTGTAAGCG 511
Qy 3389 GCACTGTCAACCGCTACGAGAGTACAGGAGCAATGACATGACCTCGGCGCGGCTAGAC 3448
Db 510 GCGCCGCTGACAGCGAGCGGAGAGTGCAGCGGAGCAATGACAGTCCGCGCGGTAGAG 451
Qy 3449 CCGACGCTGACAGGAGGCGGCTCCCGCTGACGCGACCTCCAAACGCGCGCGCG 3508
Db 450 CCGAGCTGTAAGGAGAGACGCGCTTCTGCGAGCGCGCGCTGAGAGCGCGCGCTGT 391
Qy 3509 TTGACGCGCTGATGCTGTTGCTCAAGTACTTTCGCGCGCGCGCTGCTCAACCGCGT 3568
Db 390 TCGAGGCGGAGAGCGGCTGTTCTTCAAGTACGCGCGCGCGCGCGCTGACCGCGT 331
Qy 3569 CCGAGCTGCGGAGTGCAGTGAACAGGTGACAGAGCTGCGCGAACCGCGCGCTGACAG 3628
Db 330 CCGAGCTCCGCGACCTGCGCGGACAGGCGAGGAGCGCGGAGAGCGCTCCCGTCAAG 271
Qy 3629 TCGAACGAGGCGGACATCTGCTCAACGCGGCTGAACGCGAGCGCGCTGATGAGCGGT 3688
Db 270 CCGAGCTGAGCGGCGCTCTTCAAGCGGCTGAAGGCGCTGTTGCGCTGACGACGCG 211
Qy 3689 AGTTGCGCGAGCGCGACCGGCTGCTTCCGCGGCTGTCGATGCGGACCGGTGAGCG 3748
Db 210 AGCGGACGAGAGTGGCGGCGGCTGTTCCGCGCTGTCGCGGCGGACGAGCGCGCG 151
Qy 3749 GTGACTTGGCGCTGAGCGCGCGCGACCAAGCGCGCGCGGTGCGGTGCGGTGCGCG 3808
Db 150 GCGAGGCTGTCGCGCGAGCGGCTGAGACGCGCGCGCGCGCGCTGCGGAGTGCAG 91
Qy 3809 TCGGAGTCTCCGTCGCTGACGCGCGGTGAGGAGTGTCTTCATGAC 3859
Db 90 CGTTGCGCGGCGGCGGTGCGGAGAGGCGGCGCGGCTCATCGGAGCGCT 40

```

RESULT 12  
AA256006/c  
ID AA256006 standard; DNA; 1565 BP.  
XX  
AC AA256006;  
XX  
DT 23-MAR-2000 (first entry)  
XX  
DB Contig 005 from cosmid pKOS023-27 from *Streptomyces venezuelae*.  
XX  
KW Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; contig 002;  
XX ketolide; PICCV; plicromycin; antibiotic production; narbomycin; de.  
XX  
OS *Streptomyces venezuelae*.  
XX  
FH Key Location/Qualifiers  
FT CDS 49..1507  
FT /tag= A  
FT /product= "PICCV"  
XX  
FN W09961599-A2.  
XX  
PD 02-DEC-1999.  
XX  
PF 27-MAY-1999; 99WO-US011814.  
XX  
PR 28-MAY-1998; 98US-0087080P.  
PR 28-AUG-1998; 98US-00141908.  
PR 22-SEP-1998; 98US-0100880P.  
PR 08-FEB-1999; 99US-0119139P.  
XX  
PA (KOSA-) KOSAN BIOSCIENCES INC.  
XX  
PI Ashley G, Belkacem MC, Belkacem M, McDaniel R, Tang L;

DR WPI, 2000-072618/06.  
DR P-PSDB; AAY67210.

PT New recombinant DNA encoding a domain of narbonolide polyketide synthase,  
PT for production of ketolide antibiotics.

XX Disclosure; Page 38-39; 98pp; English.

XX This is config 005 from the recombinant cosmid pKOS023-27 DNA sequence  
XX (see AAZ66001) which contains a Streptomyces venezuelae DNA insert. The  
XX cosmid contains open reading frames which encode the various modules of  
XX the narbonolide polyketide synthase (PKS). The invention relates to  
XX recombinant DNA containing a coding sequence for a narbonolide PKS.  
XX Polyketides are compounds synthesised from 2-carbon units through a  
XX series of condensations and subsequent modifications. Modular PKSs are  
XX responsible for the production of many antibiotics including picroxylin.  
XX The narbonolide PKS consists of a loading module, six extender modules,  
XX and two thioester domains. Four proteins make up the narbonolide PKS  
XX (PICAI, PICAI1, PICAI11 and PICAI1V). PICAI includes the loading module  
XX and extender modules 1 and 2, PICAI11 includes extender modules 3 and 4,  
XX PICAI1V includes extender module 5 and PICAI1V includes extender module 6  
XX and a type II thioesterase domain. The second type II thioesterase domain  
XX is found on the PICB protein. The nucleotide sequences encoding all of  
XX these proteins can be isolated in recombinant form from the recombinant  
XX cosmid pKOS023-27. Narbonolide is desosaminylated in *S. venezuelae* to  
XX yield narboxycin, the desosaminyl transferase enzyme is required for this  
XX conversion, and the desosaminyl transferase genes are also found in  
XX cosmid pKOS023-27. The recombinant DNA of the invention is used to  
XX express, in transformed cells, narbonolide (or its derivatives) or other  
XX ketolides (particularly hybrids), which may then be converted (e.g. by  
XX other enzymes recombinantly expressed in the same hosts) to polyketide  
XX antibiotics or their intermediates. The antibiotics are useful in human  
XX or veterinary medicine

XX Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;

XX Query Match 8.2%; Score 729.4; DB 3; Length 1565;  
XX Best Local Similarity 68.1%; Pred. No. 7.5e-91;  
XX Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2369 GGGCGGCTGACGCGCTTCAAGCGCGGATCAGGCGCCGATGCTGATCCACCGCTCG 2428  
DB 1530 GGGGGGATCGGGGCGGGTGGCGGTCAGCGCAGAGAGCGCGGGCTCCTCCAGCGCTCC 1471  
QY 2429 GCGACGCTCCGCTTCAAGTGTGAGCGCGCGCTCAACGATGCTGAAACCGTCAAG 2488  
DB 1470 GCGCGGCTCGGCTTCAAGTGTGAGCGCGCGCTCAACGATGCTGAAACCGTCAAG 1411  
QY 2489 AAGAACTCGTCCCGCGCTCGGATGCTGCGCGCGCGATGCTGAAAGTGTGAGCG 2548  
DB 1410 AAGTACTCGTCCCGCTTCAAGCGCGCGCGCTCGCGCGCGCTCGCGCGCGCTCGAG 1351  
QY 2549 ACCGAGTCAAGCGCTCGGCTCGGCGGTGACCGCGCGCGCGATGCTGAGCGCGCGCG 2608  
DB 1350 ACCTCGCGTGAAGGAGGTGCGGGGCTCAAGCGCGCGCGCGATGAGCGCGCGCTCC 1291  
QY 2609 AGCCCGGAGAAACCGCGCTCCCGGATGAGTGAAGCTCTCCAGCGAGTGTGAGCG 2668  
DB 1290 AGGTGCGGAGAAACCGCGCTCGGATGAGTGAAGCTCTCCAGCGAGTGTGAGCG 1231  
QY 2669 GCCACCTGCGGATGAGGCGGTGCGGCGCGATGCTTCCGCGCGCGATGCGCGAAGCTGG 2728  
DB 1230 GCGACCTGCGGATGAGGCGGTGCGGCGCGATGCTTCCGCGCGCGATGCGCGAAGCTGG 1171  
QY 2729 TCGACACCGCGAGCGAGCTCTCCAAAGCGTGAACCGAGTCCGCTGCGAGTCCCGG 2788  
DB 1170 TCGGCGCGCGTGAAGCGAGCTCTCCAAAGCGTGAACCGAGTCCGCTGCGAGTCC 1111  
QY 2789 CGCTCGCGCGGATGAGTGAAGCGAGCGAGCGCGCTCCCGCGAGTCCCGGCTCGCG 2848  
DB 1110 CGCTCGCGCGGATGAGTGAAGCGAGCGAGCGCGCTCCCGCGAGTCCCGGCTCGCG 1051  
QY 2849 TCGGACAACTGCGCGGTGCTCCGACCGCTGAGTCTCGCGGAGCGGTGAAGTGAAGC 2908

DB 1050 TGGCGAGCTTGCCTGCTCAAGCGCGCTGTAGCTCTCGGAGATGTTGACGAGTCCATC 991  
QY 2909 GCGCGGCTGCGGCTGAGCTGTTGATGCTGCGGATGAGTGAAGTGAAGTGAAGTGAAG 2968  
DB 990 GTCCTGCTGCGCGCGGCTGTTGAGGTGCGGATGAGTGAAGTGAAGTGAAGTGAAG 931  
QY 2969 TGGCGCGCGCGCGGATGAGTGTGAGGCGCGGCGCGGATGCGCGGCTGCGCGGCTG 3028  
DB 930 GAGCGAGCGCGCGGATGAGTGTGAGGCGCGGCGCGGATGCGCGGCTGCGCGGCTG 871  
QY 3029 GCGCGATCCGCGAGAGTCCCGGATGCTTCTGTCGACCGGCTGCGAGCGCGCGCTG 3088  
DB 870 GCGCGAGCTGCTGAGAGCGCGCGGATGCTTCTGTCGACCGGCTGCGAGCGCGCGCT 811  
QY 3089 CGGCTGCGGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3148  
DB 810 CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751  
QY 3149 AGGTCCGACGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3208  
DB 750 AGGCGCGAGCGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 691  
QY 3209 TAGACGCTCAAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268  
DB 690 TAGACGCTGCG 631  
QY 3269 TTGCTCAAGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3328  
DB 630 TTGCTCAAGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 571  
QY 3329 TCGTCAAGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3388  
DB 570 TCGTCAAGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511  
QY 3389 GCACTGCTCAAGCG 3448  
DB 510 GCGCGGCTCAAGCG 451  
QY 3449 CGGAGCTTCAAGCG 3508  
DB 450 CGGAGCTTCAAGCG 391  
QY 3509 TTGAGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3568  
DB 390 TTGAGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 331  
QY 3569 CGGAGCTTCAAGCG 3628  
DB 330 CGGAGCTTCAAGCG 271  
QY 3629 TCGAACCGAGCGCGCGCTTCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3688  
DB 270 CGGAGCTTCAAGCG 211  
QY 3689 AGTTCGCGAGCG 3748  
DB 210 AGCGGAGCGAGCG 151  
QY 3749 GTGACCTGCG 3808  
DB 150 GCGAGGCTTCAAGCG 91  
QY 3809 TCGCGATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3859  
DB 90 CGTTCG 40

RESULT 13  
ADA09423/c  
ID ADA09423 standard; DNA; 1565 BP.  
XX  
AC ADA09423;

XX 06-NOV-2003 (first entry)  
XX S. venezuelae DNA encoding PICCV.  
DE Streptomyces venezuelae; Sv; narbonolide polyketide synthase; PKS;  
XX narbonolide PKS; narbomycin biosynthesis; picromycin biosynthesis;  
KM PKS gene cluster; picromycin; glycosylation; hydroxylation;  
KM C12 hydroxylase; Pick; desosamine biosynthesis;  
KM desosaminyl transferase enzyme; antibiotic; narbonolide synthase; PICCV,  
KM gene, ds.  
XX Streptomyces venezuelae.  
XX OS US6509455-B1.  
XX PN 21-JAN-2003.  
XX 07-SEP-2000; 2000US-00657440.  
XX 30-APR-1997; 97US-00846247.  
PR 06-MAY-1998; 98US-00073538.  
PR 28-MAY-1998; 98US-0087080P.  
PR 28-AUG-1998; 98US-0014190P.  
PR 22-SEP-1998; 98US-0100880P.  
PR 08-FEB-1999; 99US-0139139P.  
PR 20-MAY-1999; 99US-0134990P.  
PR 27-MAY-1999; 99US-00320878.  
XX (KOSA-) KOSAN BIOSCIENCES INC.  
XX PA Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;  
PI WPI; 2003-352291/33.  
XX Novel recombinant DNA compounds comprising coding sequences for  
PT desosamine transferase gene of Streptomyces venezuelae, useful for  
PT producing desosamine transferase which transfers desosamine to substrate  
PT polyketides.  
XX Discloure; Col 75-78; 132pp; English.  
XX The present invention relates to recombinant DNA compounds that encode  
CC Streptomyces venezuelae (Sv) narbonolide polyketide synthases (PKSs). The  
CC recombinant PKSs are derived from narbonolide PKS and other genes  
CC involved in narbomycin and picromycin biosynthesis in recombinant host  
CC cells. The invention also discloses the S. venezuelae PKS gene cluster  
CC that results in the production of picromycin. Also disclosed are enzymes  
CC such as those responsible for glycosylation and hydroxylation, (e.g. C12  
CC hydroxylase (Pick)), desosamine biosynthesis, and desosaminyl transferase  
CC hydroxylase (Pick)).  
CC polyketides are useful as antibiotics and as intermediates in the  
CC synthesis of compounds for pharmaceutical applications. The present  
CC sequence encodes S. venezuelae PICCV.  
XX Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;  
SQ  
Query Match 8.2%; Score 729.4; DB 8; Length 1565;  
Best Local Similarity 68.1%; Pred. No. 7.5e-91;  
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;

QY 2549 ACCGATGTCAGAGCTCCGGTCCGAGGTGACCCGAGCCGACGTAAGCGGTGGCCCGCC 2608  
DB 1350 ACCCTGGGTGAGGAGAGTGTGGAGGGGTCAACGAGCCCGCATGTATAGCGGGTCCGCTCC 1291  
QY 2609 AGCCCGGAGAAACCGGCTCCCGGTACAGGTATACGCTCCGAGCAGGTGATCTGACAC 2668  
DB 1290 AGGTCCGAGAAAGCCGCTCGAGTACAGGTATACGCTCCGAGAGATGACCTTGACAC 1231  
QY 2669 GCCACTGCGAGATGAGCGGTCCGAGCGCATGTTTCCGAGCGGATCCGCAACAGCTGGAGC 2728  
DB 1230 GCCACTGCGAGGTCCGAGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTG 1171  
QY 2729 TCACACCCGACGACGCTCTCCACGCTTACCGAGTCCGAGTCCGAGTCCGAGTCCGAG 2788  
DB 1170 TCAGCGCCGAGTCCGAGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTG 1111  
QY 2789 CGCTCCGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTG 2848  
DB 1110 CGCTCCGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTG 1051  
QY 2849 TCAGCAACCTGCGCTGCTCCGACGCTGTAAGTCTCCGAGCGAGTGAAGTGAAGTGAAGC 2908  
DB 1050 TCAGCAACCTGCGCTGCTCCGACGCTGTAAGTCTCCGAGCGAGTGAAGTGAAGTGAAGC 991  
QY 2909 GACCGAGTCCGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 2968  
DB 990 GTCTGCGCTGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 931  
QY 2969 TTGCGCGCGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 3028  
DB 930 GAGCGAGCGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 871  
QY 3029 GCGCGCATCCGACGAGAGTCCCGAGGTCTTCTGTAAGCGGTGAGCGGTGAGCGGTGAG 3088  
DB 870 GCGCGCATCCGACGAGAGTCCCGAGGTCTTCTGTAAGCGGTGAGCGGTGAGCGGTGAG 811  
QY 3089 CCGGTGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCC 3148  
DB 810 CCGGTGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCC 751  
QY 3149 AGGTCCACAGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 3208  
DB 750 AGGTCCACAGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 691  
QY 3209 TAGACGATGAGCGAGAGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 3268  
DB 690 TAGACGATGAGCGAGAGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 631  
QY 3269 TTGATCAACGTTCCAGCGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 3328  
DB 630 TTGATCAACGTTCCAGCGCGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGT 571  
QY 3329 TCGTGTATGACCGAGTGAAGATCTCGTTGCGGAGCACACCGACTCCGCTCCGACTCTG 3388  
DB 570 TCGTGTATGACCGAGTGAAGATCTCGTTGCGGAGCACACCGACTCCGCTCCGACTCTG 511  
QY 3389 GCACCTGTCACCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 3448  
DB 510 GCACCTGTCACCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 451  
QY 3449 CCGACGCTGTACCGAGAGCGCGGTCCCGCGGTACCGAGCGGTCCGACGCGCGCGCGCG 3508  
DB 450 CCGACGCTGTACCGAGAGCGCGGTCCCGCGGTACCGAGCGGTCCGACGCGCGCGCGCG 391  
QY 3509 TTACAGCGCTGTATGTTGCTCTCAAGTCTTGGCCCGCGCGGTCTGCTCAACCGCGCTG 3568  
DB 390 TTACAGCGCTGTATGTTGCTCTCAAGTCTTGGCCCGCGCGGTCTGCTCAACCGCGCTG 331  
QY 3569 CCGACGCTGTACCGAGAGCGCGGTCCCGCGGTACCGAGCGGTCCGACGCGCGCGCGCG 3628  
DB 330 CCGACGCTGTACCGAGAGCGCGGTCCCGCGGTACCGAGCGGTCCGACGCGCGCGCGCG 271  
QY 3629 TCAGAACGACGCGCTGCTCCACCGAGGTGAAACGAGCGCGCTGATGAGCGGTG 3688



Db 270 CCCAGGTCGTGGCGGGCTCTCCAGCGGGGTGAAGGGGCTGTTCCTAGCCGACCGGCG 211  
Qy 3689 AGTTGGCGGAGCCCGGACCGCTCTGCTTCGGCGGCTGTCGATGCGGACCGGTAGGCGG 3748  
Db 210 AGCCGAGCGAGGTGGCGGGCGGTCTTCGGCTCTGTCGGCGGCGACGAGGCCCGCGCG 151  
Qy 3749 GTGACCTCGGCGCTGAGCGCCCGGACCGGCGCGGCTGCGGCTGCGGCTGCGGCGG 3808  
Db 150 GCGAGGCTCTGCGCGGACCGGCTGACCGCGCGCCCGGAGATCGGCTCGCGGGTCCGCGAG 91  
Qy 3809 TCCGCGATCTCCGTGCTACCGGCGGTGCGCGGTAGAGTGTCTTATCGAC 3859  
Db 90 CGTTCCGCGCGGGCGGTGCGGAGAGGCGGCGGCGGTCAATCGGAGCGTC 40

RESULT 14  
ADH53467/c  
ID ADH53467 standard; DNA; 1565 BP.  
XX  
AC ADH53467;  
XX  
DT 25-MAR-2004 (first entry)  
XX  
DE 5. venezuelae PK08023-26 cosmid DNA contig 005.  
XX  
KM Narbonolide polyketide synthase; PKS; desosamine;  
KM desosaminyl transferase; beta-glucosidase; pick hydrolase; antibiotic;  
KM pharmaceutical; gene therapy; antibacterial; infection; gene; ds.  
XX  
OS Streptomyces venezuelae.  
XX  
FH Key Location/Qualifiers  
FT CDS 50..1507  
FT /tag= a  
FT /product= "PICCV protein"  
XX  
PN US2003162262-A1.  
XX  
PD 28-AUG-2003.  
XX  
PF 29-MAY-2002; 2002US-00160539.  
XX  
PR 28-MAY-1998; 98US-0087080P.  
PR 28-AUG-1998; 98US-00141908.  
PR 22-SEP-1998; 98US-0100880P.  
PR 08-FEB-1999; 99US-0119139P.  
PR 20-MAY-1999; 99US-0134990P.  
PR 27-MAY-1999; 99US-00320878.  
PR 07-SEP-2000; 2000US-00657440.  
XX  
PA (ASHL/) ASHLEY G.  
PA (BETL/) BETLACH M. C.  
PA (BETL/) BETLACH M.  
PA (MCDA/) MCDANIEL R.  
PA (TANG/) TANG L.  
XX  
PI Ashley G, Betlach MC, Betlach M, Mcdaniel R, Tang L;  
XX  
DR MPI; 2003-897933/82.  
DR P-PDB; ADH53453.  
XX  
PT New recombinant DNA compound encoding a narbonolide polyketide synthase,  
PT useful for expressing recombinant polyketide synthase genes in host cells  
PT for the production of narbonolide and polyketides useful as antibiotics.  
XX  
PS Disclosure; SEQ ID NO 24; 136pp; English.  
XX  
CC The present invention relates to novel recombinant DNA compound that  
CC encodes a domain of a narbonolide polyketide synthase (PKS), a desosamine  
CC biosynthetic gene, a desosaminyl transferase gene or a beta-glucosidase  
CC gene of Streptomyces venezuelae, or a pick hydrolase gene of S.  
CC venezuelae. The recombinant DNA compounds are useful in expressing

CC recombinant polyketide synthase genes in host cells for the production of  
CC narbonolide or its derivatives and polyketides that are useful as  
CC antibiotics and as intermediates in the synthesis of compounds with  
CC pharmaceutical value. They may be used for treating e.g. bacterial  
CC infections. The invention is also useful in gene therapy. The present  
CC sequence is Streptomyces venezuelae PK08023-26 cosmid DNA 005 contig used  
CC in the invention.  
XX  
SQ Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;  
Qy 2369 GGGCGGATCGACGCGCTTCAGAGCGCGGATCAGGCGCCGATGCTATCCACCGTCG 2428  
Db 1530 GGGGGGATCGGGCGGGGTGCGGTCAAGCGAGAAAGCGGGCGCTCTCCAGCCGCTCC 1471  
Qy 2429 GCGACGTCCTGTCAGTTGATTGAGCCGGGCGGTCAACGACTGTGCGAAACGTCGAGG 2488  
Db 1470 GCGGCTTCGCGCTCAAGCTGTTCAGCGGCGGTGAGACCTGATGAAAGCTCCATG 1411  
Qy 2489 AAGAACTGTCCTCCCGCTGCGATGATCTGCGCGCCGACGTCAGCAAGTCGTGAGC 2548  
Db 1410 AAGTACTGTCGCGCTGCAAGCGCGCCCACTGCGCGCGCTGCAAGAACTCCCTACG 1351  
Qy 2549 ACCGAGTCAAGCTCCGCTCCGGGTGACCCGCGCGGACGTCGAGCGGTGCGCCGCC 2608  
Db 1350 ACTTCGCTGAGGAGGTGTCGGGGGTCAACCGCGCGCGATGATGAGCGGTGCGCGCTCC 1291  
Qy 2609 AGCCGGGGGAAACCGGCTCCCGGTACAGTACGCTTCGCGAGCTGATCTGAC 2668  
Db 1290 AGGTGCGGAGACCGGCTCGCGGTACAGTACGCTGCGCGAGATGACCTTGAC 1231  
Qy 2669 GCCACTGCGATGAGCGGTGCGGCGCATGATTCGCGCGGATCCGCAACGTCGAGCG 2728  
Db 1230 GCGACCTGCGGAGTGGCGGCGGTGCGGCGCATGATGCGGCGCTTATCCGACATTCGCG 1171  
Qy 2729 TCGACACCCCGACGAGCTCTTCAACGCGTAAACGAGTGGTCTGATGCCCGGGTTC 2788  
Db 1170 TCGGCCCCGCTGCGAGGCTGTTCAAGGCGTACCGTATGATGATGATGATGATGATG 1111  
Qy 2789 CGCTCGCGCGCTGATGACGAAACCGGCGAGGCGCTCCCGCAACTGCTGCTTCGCC 2848  
Db 1110 CGCTCGCGACCGCTCTCTGAAAGCGTTGAGGCGCTCTGAGCTGCGCGCTCTCTCC 1051  
Qy 2849 TCGCAACACTGCGCTGCTCCGACCGCTGTATGCTTCGCGGAGCGGTGACGAGTCAAGC 2908  
Db 1050 TCGGAGAGCTGCGCTGCTCAAGCGCGCTGTATGCTCTCGGAAATGTTGACGAACTGATC 991  
Qy 2909 GAGCGGTGCGGCTGAGCTGTTGATGCTGCGGAGTGAAGTGAACGAGTGTGATGAGCGG 2968  
Db 990 GTCTCTCCCTGCGCGGCGCTGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGCGG 931  
Qy 2969 TTGGCCCCCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028  
Db 930 GAGGCAAGCGCGCGGAGGACGATGAGGCGAAAGCGGATGATGATGATGATGATGATGATG 871  
Qy 3029 GCGCGCATCGACGAGAGTCCCGAGAGTCTTCTCTGACCGCTTGAAGGCGGACGCTTGG 3088  
Db 870 GCGCGAGCTGCTGAGAGGCGGCGAGTCTTCTGCGGACCGCGGAGGAGGCGGCTTCTTG 811  
Qy 3089 CCGGTGTGCGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3148  
Db 810 CCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 751  
Qy 3149 AGGTCCCAAGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3208  
Db 750 AGGCCCCAAGAGCGCGGCTGCGGCGGCTTCAAGGTCGCTGCTGAGCGCGAAGAGTTCG 691  
Qy 3209 TAGACGCTGAGGCGGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3268  
Db 690 TAGACGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 631

QY 3269 TTGGTCAACGGTTCCAGACCGCGGAGAGGATACATCGCCGTCGGGTTCTCCGACCGGACC 3328  
DB 630 TTGGTGAAGCGCTCCAGCGCGCGGAGAAATATATCCCGAGGGGTTCCCGCGGGATATC 571  
QY 3329 TCGTCAATGACCGAGTGAAGATCTCGTTGCGCGGACCAACGATCCGCGCTGACTTG 3388  
DB 570 TCGTCAATGACCGAGAACATGCGCTTGGCGCGCTCGAGGGCGGACCGGCTGTACCGG 511  
QY 3389 GCACCTGTACCCGTACCGCAAGATGACAGCGGAACTGACAGTCGTGGGCGCGGGTAAAC 3448  
DB 510 GCGCGGTGACACCGAGCGCAAGATGACAGCGGAACTGACAGTCGTGGGCGCGGGTAAAC 451  
QY 3449 CCGAGCGTGAACCGGAAAGCGGGCTCCCGCGTACCGGACCTTCAACCGCGCGCGGCG 3508  
DB 450 CCGAGCGTGAACCGGAAAGCGGGCTTCTGCGGACCGCGCTGTCGAAAGACCGCGCTGT 391  
QY 3509 TTGACCGCGCTGATCGTGTGCTCCAGTACTTTCGCGCGCGGCGCTGCTCAACCGCGTG 3568  
DB 390 TCGAGCGGAGACAGGGTGTCTTCCAGTACCGCCCGCGGGCGCGGTCTGACCGCGGTG 331  
QY 3569 CCGAGCTGCGGGATCGATGCAACAGTCCAGACAGCTCGCGAACCGCGCGCGTCCAG 3628  
DB 330 CCGAGCTGCGGGATCGATGCAACAGTCCAGACAGCTCGCGAACCGCGCGCGTCCAG 271  
QY 3629 TCGAACCGAGCGGATCGCTCCAGCGGGGTGAACCGCGCGGCGGTAGTGGCGGTG 3688  
DB 270 CCGAGCTGCGGGATCGCTCCAGCGGGGTGAACCGCGCGGCTGTTGCTGACCGACCGG 211  
QY 3689 AGTTCGCGGACCGGACCGCGCTGCTGCTGCGCGGTGTCGATGCGACCGCGTGAAGCG 3748  
DB 210 AGCGGACGAGGTGCGCGCGGTGCTGCGCGCTGCTGCGCGGACGAGCGCGCGCG 151  
QY 3749 GTGACCTGCGCGTGAAGCGCGGACCAAGCGCGCGGTGCGCGGTGCGGTGCGCG 3808  
DB 150 GCGAGGTGCTGCGCGGACCGCGGTGACCGCGCGCGGTGCGCGGTGCGCGGTGCG 91  
QY 3809 TCCCGATCTCCGTCGATGACCGCGGTGCGCGGTGAGAGTCTTCATGAC 3859  
DB 90 CGTTCGCGCGGCGGTGCGCGGTGAGAGGCGCGGTGCGGTGCGGTGCGGTGCG 40

RESULT 15  
ABSS6095/c  
ID ABSS6095 standard; DNA; 1565 BP.  
AC  
XX ABSS6095;  
XX  
XX  
XX 21-JAN-2003 (first entry)  
XX  
XX  
XX Contig 005 from cosmid pKOS023-27S containing S. venezuelae DNA insert.  
XX  
XX  
XX Narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;  
XX desosaminyl transferase gene; beta-glucosidase gene; antibiotic;  
XX pick hydroxylase gene; C12 hydroxylase gene; narbonolide;  
XX desosaminylated polyketide; narbomycin biosynthesis; mutant;  
XX plectomycin biosynthesis; ds.  
XX  
XX Streptomyces venezuelae.  
XX Synthetic.  
XX  
XX  
XX MO200297062-A2.  
XX  
XX  
XX PD 05-DEC-2002.  
XX  
XX  
XX 22-FEB-2002; 2002WO-US005642.  
XX  
XX  
XX 22-FEB-2001; 2001US-00793708.  
XX  
XX  
XX (KOSA-) KOSAN BIOSCIENCES INC.  
XX  
XX  
XX Ashley G, Belach MC, Belach M, Medaniel R, Tang L;

DR WPI; 2003-041412/03.  
XX  
XX Preparation of polyketides by recombinant DNA technology, useful as  
PT antibiotics and as intermediates in the synthesis of pharmaceutical  
PT compounds.  
XX  
XX Disclosure; Page 43-44; 127p; English.  
XX  
XX  
XX The present invention relates to recombinant DNA sequences encoding for a  
CC narbonolide polyketide synthase (PKS) domain, and methods of producing  
CC polyketides by recombinant DNA technology. The recombinant DNA sequences  
CC are derived from Streptomyces venezuelae desosamine biosynthetic,  
CC desosaminyl transferase, beta-glucosidase, or pick (C12) hydroxylase  
CC genes. The method is useful for transforming a cell with a recombinant  
CC expression vector that encodes a functional beta-glucosidase gene, and  
CC therefore for increasing the yield of a desosaminylated polyketide in a  
CC cell. The recombinant methods and materials are useful for expressing  
CC polyketides with significant antibiotic activity, derived in whole or in  
CC part from the narbonolide PKS gene, and other genes involved in  
CC narbomycin and plectomycin biosynthesis in recombinant host cells. The  
CC present sequence represents contig 005 from cosmid pKOS023-27S containing  
CC S. venezuelae DNA  
XX  
XX  
SQ Sequence 1565 BP; 225 A; 669 C; 476 G; 195 T; 0 U; 0 Other;  
Query Match 8.2%; Score 729.4; DB 10; Length 1565;  
Best Local Similarity 68.1%; Pred. No. 7.5e-91;  
Matches 1015; Conservative 0; Mismatches 476; Indels 0; Gaps 0;  
QY 2369 GGGCGGATCGACCGCGCTGACAGCGCGCGGATACAGGCGCGCGATCTATCCACCGCG 2428  
DB 1530 GGGCGGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1471  
QY 2429 GCGACGTCGCGCTTCCAGTGGTGAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 2488  
DB 1470 GCGCGGTCGCGCTTCCAGTGGTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1411  
QY 2489 AAGAACTGTCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2548  
DB 1410 AAGTACTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1351  
QY 2549 ACCGATGACGCGCTTCCAGTGGTGAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2608  
DB 1350 ACTCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1291  
QY 2609 AGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2668  
DB 1290 AGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1231  
QY 2669 GCGACCTGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2728  
DB 1230 GCGACCTGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1171  
QY 2729 TCGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2788  
DB 1170 TCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1111  
QY 2789 CGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2848  
DB 1110 CGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1051  
QY 2849 TCGGACGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2908  
DB 1050 TCGGACGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 991  
QY 2909 GCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2968  
DB 990 GTCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 931  
QY 2969 TTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3028  
DB 930 GAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 871





**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:24:14 ; Search time 43615.7 Seconds  
(without alignments)  
11760.789 Million cell updates/sec

Title: US-10-611-442-1  
Perfect score: 8942  
Sequence: 1 ggcgcgttcgacatccatcg.....gtgcgagcagcggagatcc 9024

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenBank! :  
1: gb\_ba :  
2: gb\_in :  
3: gb\_env :  
4: gb\_ov :  
5: gb\_ov :  
6: gb\_pat :  
7: gb\_ph :  
8: gb\_pr :  
9: gb\_to :  
10: gb\_str :  
11: gb\_sy :  
12: gb\_un :  
13: gb\_vl :  
14: gb\_hlg :  
15: gb\_pl :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	3037.4	34.0	8160	1	SEERYGENE
C 2	3037.4	34.0	8160	6	BD073332
C 3	3037.4	34.0	8160	6	AX000487
C 4	3032.6	33.9	8051	6	AR092722
C 5	3032.6	33.9	8340	1	SEU77459
C 6	2606.8	29.2	47981	1	AF263245
C 7	2606.8	29.2	47981	6	AX112026
C 8	1922.2	21.5	61845	1	AY623658
C 9	1513.6	16.9	210614	1	AB088224
C 10	1313.8	14.7	17665	1	AF521878
C 11	1309.4	14.6	12441	1	AF079762
C 12	1309.4	14.6	12441	6	BD217172
C 13	1309.4	14.6	58343	1	AB089954
C 14	1276	14.3	9523	1	AF237894S1
C 15	1269.8	14.2	15052	1	AF055579
C 16	1241.2	13.9	60196	6	AX697977
C 17	913.8	10.2	1206	6	BD073333
C 18	913.8	10.2	1206	6	AX000494

19	884	9.9	8407	1	SEU82823
C 20	729.4	8.2	1565	6	BD232539
C 21	729.4	8.2	1565	6	AR271685
C 22	729.4	8.2	1565	6	AR277669
C 23	729.4	8.2	1565	6	AR679279
C 24	722	8.1	1458	6	BD217175
C 25	720.6	8.1	85915	1	AY509120
C 26	712.2	8.0	3439	1	SEV14332
C 27	712.2	8.0	3439	6	BD073330
C 28	712.2	8.0	3439	6	AX000482
C 29	703	7.9	3469	1	SEU77454
C 30	701.4	7.8	3756	6	AR092721
C 31	673.2	7.5	61845	1	AY623658
C 32	667.2	7.5	1467	6	AX698013
C 33	596.6	6.7	1227	6	AX698011
C 34	595.4	6.7	1248	6	BD217174
C 35	564.8	6.3	109528	1	AF040570
C 36	547	6.1	9523	1	AF237894S1
C 37	545.6	6.1	39428	1	AF322256
C 38	545.4	6.1	15052	1	AF055579
C 39	540.8	6.0	34869	1	AF324838
C 40	534.4	6.0	18977	1	SAR7932
C 41	519.2	5.8	5970	6	BD232536
C 42	519.2	5.8	5970	6	AR271682
C 43	519.2	5.8	5970	6	AR277666
C 44	519.2	5.8	5970	6	AR679276
C 45	500.4	5.6	45055	6	AX574197

## ALIGNMENTS

RESULT 1  
SEERYGENE/LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

SEERYGENE  
S.erythraea erythromycin gene cluster DNA.  
Y11199  
Y11199.1 GI:2665334  
2,3 dehydratase; aminotransferase; dehydratase; deoxyhexose  
reductase; eryBIV gene; eryBV gene; eryBVI gene; eryBVII gene;  
eryCIV gene; eryCV gene; eryCVI gene; erythromycin;  
glycosyltransferase; TDP-4-keto-6-deoxyglucose 3,5 epimerase;  
TDP-4-keto-6-deoxyglucose 5 epimerase;  
TDP-N-dimethyldeoxosamine-N-methyltransferase.  
Saccharopolyspora erythraea  
Saccharopolyspora erythraea  
Bacteria; Actinobacteridae; Actinobacteriales; Saccharopolyspora.  
Pseudonocardineae; Pseudonocardiales; Saccharopolyspora.

Gaiser, S., Bohm, G.A., Cortes, J. and Leadlay, P.F.  
Analysis of seven genes from the eryA-eryK region of the  
erythromycin biosynthetic gene cluster in Saccharopolyspora  
erythraea  
Mol. Gen. Genet. 256 (3), 239-251 (1997)

2 (bases 1 to 8160)  
Leadlay, P.F.  
Direct Submision  
Submitted (12-FEB-1997) P.F. Leadlay, Department Of Biochemistry,  
University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK  
Related sequence: J05776.  
Location/Qualifiers  
1. 8160  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRRL2338"  
/db\_xref="taxon:1836"  
242..1210  
/gene="EryBIV"  
242..1210  
/gene="EryBIV"  
/function="Involved in mycarose biosynthesis"  
/codon\_start=1

[illegible][illegible]

Db 7607 GTGCGGGGTGAATCTGAAACCGCCCGCACCGCCAGCTCACTGCAACTCCGGTTC 7548  
QY 2388 AGACGGCGGGATATCAGGCGCCCGAATGCTATCCACCCGTCGGGAGTCCCGTTGAGTT 2447  
Db 7547 ACCTTCGCGCAGAGAAAGCCCGGTGCTCTCCACCCGTCGGGATGTCGTTCGTCAGCT 7488  
QY 2448 GATTGAGCCGGGCGGTCAACGACTGTGTAACCGTTCAGAGAAAGAACTGTCGCCCGGCT 2507  
Db 7487 GATTGAGCCGTGGCGGTCAACGACTGTGTAACCGTTCAGAGAAAGAACTGTCGCCCGGCGC 7428  
QY 2508 GCGGATTCGATGCTGCGCGCCCGAGTGAACGAACTGTCAGACCGAGTGAAGCTCCGGT 2567  
Db 7427 GGGGCTGCAAGCCCTGCTTCTCCAGCAGAAATGTCACACCTCCGCGAGGCTGATCG 7368  
QY 2568 CCGGAGTGAACCGGCGCGGACGATGAGCGGGTGGCCCGGACGCGGGGAAACCGGCT 2627  
Db 7367 ACGGGATGAACCGGCGCGGACGATGAGCGGGTGGCCCTCAGCTCCGGAAGCCGCT 7308  
QY 2628 CCGGATCAGAGTGAAGCTCTCCGAGCAGTCTGATCTTCACCGGCACTTCGCGGATGGGCGG 2687  
Db 7307 CCGGATCAGAGTGAAGTTCGCGAGCAGTCTGATCTTCACCGGCACTTCGCGGAGTGGCGG 7248  
QY 2688 TCGGCGCATCGTTTCGCGCGGATCCGCAACAGCTGAGCGTCAACCCCGACGAGGC 2747  
Db 7247 TGGGACGATCTCTCCGCGCGGATGCGCAGCAGCTCGGCTCAACCCCGCGAGGC 7188  
QY 2748 TCTCCACCGGATCCCAAGTCCGTCGATGACCGGGGTCCTCGCGCGGCTGATCGA 2807  
Db 7187 TCTCCAGGGGATGAGCCAGGTCGATGACCGGGGTCCTCGCGCGGCTGATCGA 7128  
QY 2808 CGAACCGGCGAGGCGCTCCGCAACTCGGTCGCTTCGCGCGGCAACCTGCGGCT 2867  
Db 7127 CGAACCGGCAACAGGCTCCGCGAGCTCTGCGCTCGAGTCCGACGCGGCGGCTGCT 7068  
QY 2868 CCGGACCGCTGATGCTCTCGCGAGCGGTGAAGTGAAGCGGCGGCTGCGGCTGAGCT 2927  
Db 7067 CCGGCGCGCTGATGCTCTCGCGAGCGGTGAAGTGAAGCGGCGGCTGCGGCTGAGCT 7008  
QY 2928 CGTTGAGTTCGCGGATGAAGTTCGACAGTTCGATGAGCGGTCGCGCGGCGGCTGAGCT 2987  
Db 7007 CGTTGAGTTCGCGGATGAAGTTCGACAGTTCGATGAGCGGTCGCGCGGCGGCTGAGCT 6948  
QY 2988 TGAATGATGAGGCGAGCGGCGGATCGGTCGCGGTCGCGGCGGCTGAGCT 3047  
Db 6947 TGAATGATGAGGCGAGCGGCGGATCGGTCGCGGTCGCGGCGGCTGAGCT 6888  
QY 3048 CCGGAGGATCTCTGACCGGTCGAGGCGGCAAGCTGTCGCGGTCGCGGTCGAGCT 3107  
Db 6887 CCGGAGGATCTCTGACCGGTCGAGGCGGCAAGCTGTCGCGGTCGCGGTCGAGCT 6828  
QY 3108 CGTCTGTTTCGAGCCGCTGACAGGAGTACGAGCCGCGCGGAGCTCCACAGGCGGCT 3167  
Db 6827 CGTCTGTTTCGAGCCGCTGAGGAGTACGAGCCGCGCGGAGCTCCACAGGCGGCT 6768  
QY 3168 GCGGCGGAGGCTCTGTTGAGGAGGCGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGG 3227  
Db 6767 GCGGAGTTCAGGCTCTCTGAGGAGGCGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGG 6708  
QY 3228 CCGGCGGAGGCGGTCGACCAACGCGCGGATACCGGTCGAGTCAACGATTCAGGCG 3287  
Db 6707 CCGGCGGAGGCGGTCGACCAACGCGCGGATACCGGTCGAGTCAACGATTCAGGCG 6648  
QY 3288 CCGCGGAGGAGTTCATGCGCGTGGGCTCTCCGAGCGGCACTGTCGATGAGCGAGTGA 3347  
Db 6647 CCGCGGAGGAGTTCATGCGCGTGGGCTCTCCGAGCGGCACTGTCGATGAGCGAGTGA 6588  
QY 3348 GGAATCTGTTGCGGCGGACCAACGATCCGCGGTCGATCTTGGCAACGTCACCGGATCG 3407  
Db 6587 GCGTCTCTGTTGCGGCGGAGCGGATCGGCTCTGAGGCGGCAACCGGTCACCGGATCG 6528  
QY 3408 AGAAGTGAAGCGGAACTATGACGTCGCGGCGGCGGTCAGGCGGATTCGAGGAGG 3467  
Db 6527 AGAAGTGAAGCGGAACTATGACGTCGCGGCGGCGGTCAGGCGGATTCGAGGAGG 6468

QY 3468 CCGGCTCCGCGGATACCGAGGCTCCAAAGCGCGGCGGCTTCAAGCGGCTGATCGGCT 3527  
Db 6467 CAGGCTTGGGATGAGCCGCGGCTCCAGTGGCTGCGGCTGAGGCGGCTTGAATCTGCT 6408  
QY 3528 TGTCTGAGTACTTGGCGCGGCGGCTGCTCAACCGCGGTCGAGCTGCGGATGCGAT 3587  
Db 6407 TGGACCAAGTACTTGGCGCGGCGGCTGCTCAACCGCGGTCGAGCTGCGGATGCGGCG 6348  
QY 3588 CGAACAGTTCAGAGCTGCGCGAACCGGCGCGGCTGCTCAAGTTCAGTGAACCGAGCATCT 3647  
Db 6347 CGAACAGTTCAGAGCTGCGCGAACCTGCGGCTGCTCAAGGCGGAGCGGCAAGCGCTCT 6288  
QY 3648 GCTCCAAAGGAGGAGGCGCGGATGAGGCGGATGAGGCGGATGAGGCGGAGCGGAGC 3707  
Db 6287 GCTCCAGGCGGAGGAGGCGGATGAGGCGGATGAGGCGGAGGCGGAGGCGGAGGCGG 6228  
QY 3708 CTTGCTGTTTCGCGGCTGCTGAGTGAAGCGGCAACCGGTCGAGGCGGCTGAGGCG 3767  
Db 6227 CTTGCAAGTTCGCGGCTGCTGAGTGAAGCGGCAACCGGTCGAGGCGGCTGAGGCG 6168  
QY 3768 CCGGCAACGAGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 3827  
Db 6167 CCGGCAACGAGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 6108  
QY 3828 CCGCGGTCGCGGTCGAGAGTCTTCAATGAGCTGAGAACCTTCTGCGGCTGCTGCTG 3887  
Db 6107 GGGCGGTCGCGGTCGAGAGTCTTCAATGAGCTGAGAACCTTCTGCGGCTGCTGCTG 6048  
QY 3888 GAGGATCAGGAACCGTTCGCTTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 3947  
Db 6047 GAGGATCAGGAACCGTTCGCTTCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 5988  
QY 3948 GTGTCGAGGAGTTCGAGTGAAGTTCGCAACCGGCGGATGCTCACTGCAACCGAGG 4007  
Db 5987 GTGTCGAGGAGTTCGAGTGAAGTTCGCAACCGGCGGATGCTCACTGCAACCGAGG 5928  
QY 4008 CCGGTCGAGGAGGCGGAGCCGCTTGGGAGGCACTCGGTCGTCGTCGAGCGGCT 4067  
Db 5927 CCGGTCGAGGAGGCGGAGCCGCTTGGGAGGCACTCGGTCGTCGTCGAGCGGCT 5868  
QY 4068 GGTCTCGGTCGAGGAGGCGGATTCGTCGAGGAGGCGGAGGAGGAGGAGGAGGAGG 4127  
Db 5867 GGTCTCGGTCGAGGAGGCGGATTCGTCGAGGAGGCGGAGGAGGAGGAGGAGGAGG 5808  
QY 4128 ACCTTCTCGGCGGAGGATTCGTCGAGGAGGCGGATTCGTCGAGGAGGAGGAGGAGG 4187  
Db 5807 ACCTTCTCGGCGGAGGATTCGTCGAGGAGGCGGATTCGTCGAGGAGGAGGAGGAGG 5748  
QY 4188 TCCACCAAGATGATCACTGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4247  
Db 5747 TCCACCAAGATGATCACTGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5688  
QY 4248 GTGAGCGGCGGATGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4307  
Db 5687 GATATTCGCGGAGCTGTCGAGGAGGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAG 5628  
QY 4308 TCCGTCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4367  
Db 5627 ACCCTGTCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5568  
QY 4368 ATCTTTCGTCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4427  
Db 5567 ATCTTTCGTCGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5508  
QY 4428 ATGACCGGATCGGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4487  
Db 5507 ATGACCGGATCGGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5448  
QY 4488 GCGGTCAGCGCTTGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4547  
Db 5447 GAGGTCAGCGCTTGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 5388

OY	4548	GGAGTCGCA	CCCGTCGTG	CAGCCCA	GGGCGCTGG	CGCGCTCGA	GAAGAAGCTTGA	CTTGG	4607								
Db	5387	GGCGCCCG	CCGCGGTG	CAGCCCA	GGCGCGTGG	CGCGCTCGA	GAAGAAGCTTGA	CTTGG	5328								
OY	4608	TGCTCGCG	GCATCTT	CGCAG	CGCTTCA	CAGTCTGTGT	CGGCCCCA	CAGTGTACG	4667								
Db	5327	TGCTCGCG	GCATCTT	CGCAG	CGCTTCA	CAGTCTGTGT	CGGCCCCA	CAGTGTACG	5268								
OY	4668	CCGACGAT	CGCGCG	GTCTG	CGGGGTGA	CGAGCGCG	CGACGTGTGT	TCGGGTGAC	CA	4727							
Db	5267	CCGATGAT	CGCGCG	GTCTG	CGGGGTGA	CGAGCGCG	CGACGTGTGT	TCGGGTGAC	CA	5208							
OY	4728	CCGGTCCG	CGGGGTG	CACTGTG	CGAACA	CAACCGGTGT	GTGTGTGT	CGAGCCAG	CTT	CGCGGTG	4787						
Db	5207	CCGGTCCG	CGGGGTG	CACTGTG	CGAACA	CAACCGGTGT	GTGTGTGT	CGAGCCAG	CTT	CGCGGTG	5148						
OY	4788	GGCGGTGG	CCGGAAG	GTCA	TTCGAG	CGGATCA	CTTCA	CCCGGTGA	CGTCA	CCGCG	4847						
Db	5147	GGCGGTGG	CCGGAAG	GTCA	TTCGAG	CGGATCA	CTTCA	CCCGGTGA	CGTCA	CCGCG	5088						
OY	4848	AGCACGAT	TCGAGCG	ACGCTG	CGGTG	CGATG	CAGTGC	CGTAC	CCCGAC	4907							
Db	5087	AGCACGAT	TCGAGCG	ACGCTG	CGGTG	CGATG	CAGTGC	CGTAC	CCCGAC	5028							
OY	4908	AGGTGCG	GCACCC	CGGCTT	CGAAT	CTCCCG	CA	CAAGGG	CCCGCG	CGTGTGTGA	GC	CAATTG	4967				
Db	5027	AGGTGCG	GCACCC	CGGCTT	CGAAT	CTCCCG	CA	CAAGGG	CCCGCG	CGTGTGTGA	GC	CAATTG	4968				
OY	4968	TTGTTCAG	GGGCCA	CTCCAG	CGGGCC	AGAA	CCGCTG	CGCGT	CGCGAT	CGT	CGG	CGG	5027				
Db	4967	TTGTTCAG	GGGCCA	CTCCAG	CGGGCC	AGAA	CCGCTG	CGCGT	CGCGAT	CGT	CGG	CGG	5028				
OY	5028	CCCA	CGTGCAG	GGGGT	GCAGAA	CGCT	CGGGAC	CGCGGA	GAT	CGCCAG	AT	CGCGT	CGT	5087			
Db	4907	CCCA	CGTGCAG	GGGGT	GCAGAA	CGCT	CGGGAC	CGCGGA	GAT	CGCCAG	AT	CGCGT	CGT	5088			
OY	5088	ACGCGCTT	CA	CGCGCT	CGCGGT	TTTGA	CGGCG	GAAC	CGCG	CAAG	CGCA	CGAG	CTT	CGCG	5147		
Db	4847	ACGCGCTT	CA	CGCGCT	CGCGGT	TTTGA	CGGCG	GAAC	CGCG	CAAG	CGCA	CGAG	CTT	CGCG	5148		
OY	5148	GCTTGA	TTTGA	CGTAA	GGCTGT	GCAG	ACGAT	CGGTAT	CTTGG	CCGAA	GGT	CA	CTT	CGCG	5207		
Db	4787	GCTTGA	TTTGA	CGTAA	GGCTGT	GCAG	ACGAT	CGGTAT	CTTGG	CCGAA	GGT	CA	CTT	CGCG	5208		
OY	5208	ACCGGTGT	CGGGG	CA	CGCGCT	CGTGA	AGT	CGCG	CGG	CACT	CGAC	GA	CTT	CGCG	5267		
Db	4727	ACCGGTGT	CGGGG	CA	CGCGCT	CGTGA	AGT	CGCG	CGG	CACT	CGAC	GA	CTT	CGCG	5268		
OY	5268	GGTTCCT	CGTTCG	GTAA	ACCG	CCCA	CCCTCT	CGA	AGT	CGCAG	CGT	CGTAA	GC	5327			
Db	4667	GGTTCCT	CGTTCG	GTAA	ACCG	CCCA	CCCTCT	CGA	AGT	CGCAG	CGT	CGTAA	GC	5328			
OY	5328	GGGTAC	CGGGAC	CGT	CGAC	ACGTAA	GTAA	GTAA	GTAA	GTAA	GTAA	GTAA	GTAA	GTAA	5387		
Db	4607	GGGTAC	CGGGAC	CGT	CGAC	ACGTAA	GTAA	GTAA	GTAA	GTAA	GTAA	GTAA	GTAA	GTAA	5388		
OY	5388	CGGTGA	GTGT	CGCGCT	GA	CACTG	CA	CGCTG	CA	CGCTG	CA	CGCTG	CA	CGCTG	5447		
Db	4547	CGGTGA	GTGT	CGCGCT	GA	CACTG	CA	CGCTG	CA	CGCTG	CA	CGCTG	CA	CGCTG	5448		
OY	5448	CGCGCT	CGGTG	CGAGT	GCAG	CGCGT	GTGA	GGGTG	CGCTG	AT	CTT	GA	CA	CGA	5507		
Db	4487	CGCGCT	CGGTG	CGAGT	GCAG	CGCGT	GTGA	GGGTG	CGCTG	AT	CTT	GA	CA	CGA	5508		
OY	5508	GGGCGAG	CA	TGCGTGT	GTGGGG	GA	GA	GA	GGT	GTGAT	CGAGA	CTT	CA	CTT	CG	5567	
Db	4427	GGGCGAG	CA	TGCGTGT	GTGGGG	GA	GA	GA	GGT	GTGAT	CGAGA	CTT	CA	CTT	CG	4368	
OY	5568	GGTGA	CTGT	CGTCA	CGAG	CGCGGA	AGAT	CTG	GA	GA	CTT	CTT	CTT	CTT	CTT	CGA	5627
Db	4367	GGTGA	CTGT	CGTCA	CGAG	CGCGGA	AGAT	CTG	GA	GA	CTT	CTT	CTT				

Db	4248	TGCCGTGTCGGCGCGGATCCACCCCGTGGCTGATGTCGGGCAAGCGGCGCGCT	4248
OY	5588	GGACGAACTCACTGTGTGTGTCGTCCGAGATCCAGCTCAAGCAGGATTTCAATGTTT	5747
Db	4247	GCAAGAACTGGCGCTGCGCGACGTCCGAGATCCAGCTCAACAGGCTGTTCATGTTGT	4188
OY	5748	GGACCGGGGGTCCCGCCCGAAGAAAGCAGCAGTCCGGATCTCGAAGACCCCTGAGG	5807
Db	4187	TGGCCGGGGGTCCCGCCCGAATGAAGATGACAGACCGGGCTGGAAGAGCCCTCGG	4128
OY	5808	GCACTGCTCCAGTACCCCGGCGGACCGCCCGGCTTCCACTGTGTGATATGAGGCCA	5867
Db	4127	GCAAGCGCGAGGACGTGTCGTGGCCCGGATCCCGCTCAACGAGCGTGGGACGAGGCCA	4068
OY	5868	GCAAGGAGGACCTGCAATTCGAGGAGTCCGAGCGGATCTCGAAGACCTTCGACCA	5927
Db	4067	GCAAGGAGGACCTGCAATTCGAGGAGTCCGAGCGGATCTCGAAGACCTTCGACCA	4008
OY	5928	GCAAGGAGGACCTGCAATTCGAGGAGTCCGAGCGGATCTCGAAGACCTTCGACCA	5987
Db	4007	GCAAGGAGGACCTGCAATTCGAGGAGTCCGAGCGGATCTCGAAGACCTTCGACCA	3948
OY	5988	TGTTGCGGATGCTGTTGCGCAGAGAACGAGAACCTCTGCTGCACTGAGGACGTGACGA	6047
Db	3947	TGTTGCGGATGCTGTTGCGCAGAGAACGAGAACCTCTGCTGCACTGAGGACGTGACGA	3888
OY	6048	GGATCCGGTGGGGCGGGTCCGTCGAGATCTGATGAACCTGACCGGATTCGTTGT	6107
Db	3887	GGATCCGGCTCGGGCGGGTCCGTTGGAATCTGATGAACCGGACCTTCGAGCGGGGT	3828
OY	6108	GGACCTTCATGTAATTCGTGCGGGTGGCCGTGAGAGGTCCGAGAGCTGGACCGGCTTGA	6167
Db	3827	GGACCGCGGTGTAAATTCGTGCGGGTGGCCGTGAGAGGTCCGAGAGCTGGACCGGCTTGA	3768
OY	6168	TGTTGCGCAGTTCGCGCTTGGCTTGCAACAAGCGTCAATACGCGGTGCACTCCGCA	6227
Db	3767	TGTTGCGCGGCTCGGCTTGGCTTGCAACAAGCGTCAATACGCGGTGCACTCCGCA	3708
OY	6228	CGATCAGACCGAGGAACCGGATCTCGGGTTGGAAGATGATGGTTGAATCCAGTCCGTC	6287
Db	3707	CGATGAGGCGAGGAAGCGGATCTCGGGCTTGCAAGATGATGGGCTGATCCAGTCCGCGC	3648
OY	6288	GCCATCCGAATTTGTCGCGACGTGACAGGCGCTTCGATGAGAGAAAGCGCCGAGTCGT	6347
Db	3647	GCCAGCCGAATTTGTCGCGACGTGACAGGCGCTTCGATGAGAGAAAGCGCCGAGTCGT	3588
OY	6348	GCGCAGGCGGACCGGCTCTCGGGGTGGAACGACGAGCTTCAGATGATGTAAGGCACTC	6407
Db	3587	GCGCGAGGCTTCGCGCTCTCGGGGTGGAACGACGAGGTCATGGCGTGAAGCGGACGC	3528
OY	6408	GATGCACTCGAGCGGATGCTCGGCGGCGCGGTGGCCCAACGATCGTGAAGTCGTCGG	6467
Db	3527	GCGCACCCTCGAGCGGCTGCTCGGCGCGGTGGCGGTCAGCAGTCGTGGGTTTCCGCG	3468
OY	6468	TGGCGGTGGGAGTTCGCGCGGTGCGGAGTCAAGAAACGATTTGCGGATTTGTGATTCGG	6527
Db	3467	TGGCGGTGGGAGTTCGCGCGGTGCGGAGTCAAGAAACGATTTGCGGATTTGTGATTCGG	3408
OY	6528	GAGTGCAGTACCGGTTGACCGGATCCCGCATACGCTCTCCGATGATGTCGTGGAGCGGTC	6587
Db	3407	GAGTGCAGGTCGCGGTCGCGGATCCGATACCATGACTCTCCGCGGTTCCTTCACCGGCT	3348
OY	6588	CGTGCATACCGCCCGGACGTGACATTCGTGATCAAGCCGCGCCAGTGAAGGCTCC	6646
Db	3347	CGGCTTGGCGCGCGCGGGC-----GTGTGATCAAGACCGCACTCACTGAAGGTCCT	3293
OY	6647	GCCCGCAGCGGAGAAAGGTC 6667	
Db	3292	GGTCAAGCGCGGAGAAAGGCGC 3272	

[illegible]

QY	2388	GAGCGGCGGGAATCAGGCGCCCGGAGCTGATCCACCCGTCGGCGACGTCGGTCAAGTT	2447
Db	7547	ACCTTCGCGCGAGAGACCGCGGTGTCCTCCACCCTGCGGCAATCTCTGTTGACACT	7488
QY	2448	GATTAGCGCGGCGGCTCAACCGACTGTGTGAAACCGTCGAGGAGAACTCGTCCCGCGCT	2507
Db	7487	GATTAGCCCGTGGCGGTCAACCGACTGTGTGAAAGCCGTGAGGAGATACTCGTGCGCGGGCG	7428
QY	2508	GCGGATCGAGTCTCCGCGCCCGACCTTGACGAAGTGTGTGAGACACCGAGTGCAGGCTCCGCT	2567
Db	7427	GGGCGTCGACGCGCTTCGTTCTTCAGGCAAGAACTTCCACACACTTCGCGAGGCTGTGTCG	7368
QY	2568	CCGCGGATGACCCGCGCGCGACCTGATAGCGGATGCGCCCGCGACCCGCGGAGAAACCGGCT	2627
Db	7367	ACGGGGATGACCCGCGCCCGCGACCTGATAGCGGATGCGCCCGCGACCCGCGGAGAAACCGGCT	7308
QY	2628	CCCGGTACAGGTATACGCTCTTCAGGCAAGTGTGATCTGACACCGCACTCTGCGGATGCGCGG	2687
Db	7307	CGCGGTATAGGTATACGCTCTTCAGGCAAGTGTGATCTGACACCGCACTCTGCGGATGCGCGG	7248
QY	2688	TGCGGCGCATGTTTCCGCGCGGATTCGCAACAGTGGGCGGTGCAACCCCGACAGCAGGC	2747
Db	7247	TGGGACGCACTGTCTTCGCGCGGATTCGCAACAGTGGGCGGTGCAACCCCGACAGCAGGC	7188
QY	2748	TCTTCAACCGCGTAAACCGAGGTGCGTCTGCAATGCCCCGCGGATCCGCTCGGCGCGGTATGCA	2807
Db	7187	TCTTCAAGGCGTAAACCGAGGTGCAATGCAATGCCCGGGATCCGCTCGGCGCGGTATGCA	7128
QY	2808	CGAACCAGGCGAGGCGCTTCGCGCACTCCGCTCGCTTGGCCCTCGGACAACTGCGCTGCT	2867
Db	7127	CGAACCAGCAACGAGCGCTTCGCGCACTCTGTTGGCTGTGAGTGTGCAAGCGCGCGCTGCT	7068
QY	2868	CCCGACCGCTGTAGTCTTCGCGGACGAGTGAAGAACTGCAAGCGCGCGGTGCGGCTGCACT	2927
Db	7067	CGCGCGCGCTGTAGTCTTCGCGGACCGTGAAGAACTGCAAGCGCGCGGTGCGGCTGCACT	7008
QY	2928	CGTTGAGTTGCGCGATGAAGTCCACCAAGTGTGATGAGCGGTGAGCCCGCGCGGACAGA	2987
Db	7007	CGTTGAGCTCGCGCATGAAGTGTGAGCGGTGAGCGCGGTGAGCCCGCGCGGACAGA	6948
QY	2988	TGATGTGATGAGGCGGAGCGGACCGGTGCGCGGTGCGGCGGACATCCGACGAGAGT	3047
Db	6947	TGATGTGATGAGGCGGAGCGGAGTGTGCGCGGTGCGCGGTGCGGCGGACATCCGACGAGAGC	6888
QY	3048	CCCGAGAGTTCTTCCTGACCCGTTCCAAAGGCGCACGCTTGCGGTGTGCGCTGTGATCT	3107
Db	6887	CTTGCAGAGTTCTTCTTGAACCGTTCCAAAGCGCGCGCTTGCGGTGTGCTGTGATCT	6828
QY	3108	CGTGTGTTGTCAGTCCCGCTACAGGAGGTATAGGACCCGCGCGAGGTCCCAACAGGCGCGGCT	3167
Db	6827	CGTGTGTTGTCAGTCCCGCTACAGGAGGTATAGGAGTGTGCGGACCTCCCAACAGGCGCGGCT	6768
QY	3168	GCGCGCGCACGCTCTGTTCGGTGAGGCGGAAAGCGTGTGTGATGACGCTCAGGCGGAAAGC	3227
Db	6767	GCGCGCTTCAACGCTGTCTCGGTGAGGCGGAAAGCGTGTGTGATGACGCTCAGGCTCAGAAAC	6708
QY	3228	CCCGCGCGGCGCGGTGCGACCAACGCGCCCGATACCCCGGATGTGTCAACGCTTCCAGGC	3287
Db	6707	CGCGCTCGGCGCGGTGCGACCAACGCTGCGCGGACCGGAGGTGTGTACGCGCTCCAGGC	6648
QY	3288	CGCGCGAGAGGTATCATTCGCGGTGAGGTTCTCCACCGGACCTGTGTGATGACCGAGGTGA	3347
Db	6647	CGCGCGAGAGGTATCATTCGCGGTTGTCCGTGTGAGGACCTGTGTGATGACCGGCGCA	6588
QY	3348	GGAATCTGTGTTCCGCGGACCAACCGACTTCGCGGTGTGATGACCTGTTCACCCGTATGCG	3407
Db	6587	GCGTCTGTGTTCCCGCGGAGACCGATGCGGCTCTGTAGCGGCGACCGGTATCACCGCACGC	6528
QY	3408	AGAAGTGAAGCGGAAATGCAAGTGTGGGCGCGGAGTGAAGCGCGGACGCTGTAGGGAGAGG	3467
Db	6527	AGAAGTGAAGCGGAAATGCAAGTGTGGCGCGGAGTGAAGCGCGGACGCTGTAGGGAGAGG	6468



Qy 3468 CCGGCTCCCGCGGATCCGACAGCTTCAACCGCGCGGCGTTGACGGCGCTTATCTGCT 3527  
Db CAGGCTTGGCGGTACACCGCGCGCTGCAATGCGCTGGCGGCTTCAAGCGCTTATCTGCT 6408  
Qy 3528 TGTCTCGAGTACTTCCCGCGCGGCGCTTCAACCGCGCTTCAAGCTTGGCGAT 3587  
Db TGGACGAGTACTTCCCGCGCGGCGCTTCAACCGCGCTTCAAGCTTGGCGAT 6348  
Qy 3588 CGAACGAGTCCGAGAGCTCCGCGGACCGCGCGCTTCAAGCTTCAAGCGGCGATCT 3647  
Db CGAACGAGTCCGAGAGCTTGGCGGACCTTCCGCGCTTCAAGCGGCGATCT 6288  
Qy 3648 GCTTCAACCGGCGTGAACCGCGAGCGCGCTTGAAGCGGCGTGAAGTTCCGCGAGCGGACCG 3707  
Db GCTTCAAGCGGCGTGAACCGGCGGCGCGCTTGAAGCGGCGGCGGCGGCGG 6228  
Qy 3708 CTTGTCTGTTCCGCGGCTGTCGATGCGGACCGGCGGCGGCGGCGGCGGCGG 3767  
Db CTTGCAAGCTTCCGCGCTTCAACCGCGCGCGGCGGCGGCGGCGGCGG 6168  
Qy 3768 CCGCGACCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3827  
Db CCGCGACCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6167  
Qy 3828 CCGCGGCTCCGCTGAGAGTGTCTTCAATCGAGCGGACCTTCTGCGCTGCTGCTG 3887  
Db GGGCGGTTTCCGCTGAGAGTGTCTTCAATCGGCGGAGCGGACCTTCTGCGCTGCTGCGGCA 6048  
Qy 3888 GAGATTCAGAACGCTTGGCTTTCGCTTTCCTTCCGCGGCTTGAACCGGCGGCTG 3947  
Db GGTGCTCAAGAACGCTTGGCGGCTTCTGCTTTCCTTTCGCGGCTTGAATCGGCGGCTG 5988  
Qy 3948 GTGGGCGGAGTCCGAGTACGTCGACACCGCGCGGAGTCTCTCACTGAACCGAGGCG 4007  
Db GTGGCGGCGAGCGGAGTATGTCGACACCGCGCGGAGTCTCTCTGAGACCGGCGGCGG 5928  
Qy 4008 CCGGTCGCGAGGCGGACCGCGTTCGCGAGCGCACTCGCTCTGCTGCGCGGCGGCGGCT 4067  
Db CCGGTCGCGAGCGGAGCACCGCGGTCGAGAGCTGTTGCTGCTGCTGCTGCGAGCGGCGG 5868  
Qy 4068 GGGCTCCGTCGAGGCGGCTCACTGTCGAGCGGCGGCGGAGAGTACGCGCTGTCGAGCG 4127  
Db GGGCTCCGTCGAGGCGGCTCACTGTCGAGCGGCGGCGGAGAGTACGCGCTGTCGAGCGCA 5808  
Qy 4128 ACGTTCTCCGCGCGGAGATCGCGCTGCGAGCGGCTCAAGTGCAGGTCGCGGCTGCGG 4187  
Db ACGTTCTCCGCTCCGAGATCGCGCTGCAACTGAGTGCAGTGCAGTGCAGTGCAGG 5748  
Qy 4188 TCCACGAGATGATCAAGTACTGATGATGCTCTCTGTCGCGCGGAGCGAGTGCAGCG 4247  
Db TCCACGAGATGATCAAGTACTGATGATGCTCTCTGTCGCGAGATTCGAGCGCGGTCAGCG 5688  
Qy 4248 GTGACGCGCGGATCGTCGCGAGCTCGCTGTCGAGCGGCGGCTGTCAGCGGCTTTCGCG 4307  
Db GATATGCGCGGATCGTCGCGAGCTCGCTGTCGAGCGGCGGCTGTCAGCGGCTTTCGCG 5628  
Qy 4308 TCCCGGCTCCGCGGAGCGGCGGTCGAGGAGGAGTGAAGCGGCGGCGGCGGCGGCGGCGG 4367  
Db ACCCTGCTCCGCGGAGCGGCGGTCGAGGAGGAGTGAAGCGGCGGCGGCGGCGGCGGCGG 5568  
Qy 4368 ATCTTCCGCTGAGTCCGATCTCGGTGACCACTTGTCCGCGCGGATGCCAGAGTTGTGC 4427  
Db ATCTTCCGCTGAGTCCGATCTCGGTGACCACTTGTGTCCGCGGATGCCAGAGTTGTGC 5508  
Qy 4428 ATGCGCCGAGATCGTTCCGCGAGCGGCGGCTGTCGAGTGAACCGGCGGCGGCGGCGGCGG 4487  
Db ATGCGCGGAGTCCGCTCCGCGAGCGGCGGCTGTCGAGTGAACCGGCGGCGGCGGCGGCGG 5448  
Qy 4488 GCGGTCGAGCGGCTTGTGCGGCGGAGTGAACCTGCGGCGGCGGCGGCGGCGGCGGCGG 4547  
Db GAGGTCGAGCGGCTTGTGCGGCGGAGTGAACCTGCGGCGGCGGCGGCGGCGGCGGCGG 5388  
Qy 4548 GAGCTTCACCGGCTGTCAGCGGCGGCGGCGGCTGCGAGAGAGAGTTCGAGCTTGC 4607

Db 5387 GGGCGCGCGCGCGCGGTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5328  
Qy 4608 TGGTCGCGCGGAGTCTTCCGCGAGCGCTTCCAGAGTCTGATCGGCGGCGGCGGCGGCGGCGG 4667  
Db TGGTCGCGCGGAGTCTTCCGAGCGCTTCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5268  
Qy 4668 CCGAGATTCGCGCGGCTTCCGCGGAGTGAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4727  
Db CCGAGATTCGCGCGGCTTCCGCGGAGTGAACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5208  
Qy 4728 CCGGTCGCGGAGTGAAGTGAACGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4787  
Db CCGGTCCTCCGAGTTCAGTTCGAGAGACCGGTTCCAGGCTTCCAGGCTTCCGCGGCGGCGG 5148  
Qy 4788 GCGGTCGCGGAGGATTCATTCGAGGATGATCACTTCAACCGGTCGATCAACCGGCGGCGG 4847  
Db GCGGTCGCGGAGGATTCATTCGAGGATGATCACTTCCGCGGACGATCGCTGCGGCGGCGG 5088  
Qy 4848 AGCACGAGTTCAGAGGCGAGCGGTCGATGAGGTCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 4907  
Db AGCACGAGTTCAGAGGCGGAGCGGTCGATGAGGTCGAGTGCAGTGCAGTGCAGTGCAGTGCAG 5028  
Qy 4908 AGGTCGCGGAGCGGCGGCTTGAATCTCCGCGACCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 4967  
Db AGGTCGCGGAGCGGCGGCTTGAATCTCCGCGACCGAGTGCAGTGCAGTGCAGTGCAGTGCAG 4968  
Qy 4968 TTGTTCAAGGCGGCTTCAAGCGGCGGCGGAGAACCGGTCGCGGTCGAGTGCAGTGCAGTGCAG 5027  
Db TTGTTCAAGCGGCTTCAAGCGGCGGCGGAGAACCGGTCGCGGTCGAGTGCAGTGCAGTGCAG 4908  
Qy 5028 CCGACGTCGAGGAGTTCAGAGAGCGCTTCCGAGCGGCGGCGGAGAGTCCGAGTGCAGTGCAGTGC 5087  
Db CCGACGTCGAGGAGTTCAGAGAGCGCTTCCGAGCGGCGGCGGAGAGTCCGAGTGCAGTGCAGTGC 4848  
Qy 4907 CCGACGTCGAGGAGTTCAGAGAGCGCTTCCGAGCGGCGGCGGAGAGTCCGAGTGCAGTGCAGTGC 4848  
Qy 5088 ACGCGCTTCAAGCGGCTGCGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5147  
Db ACGCGCTTCAAGCGGCTGCGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4788  
Qy 4847 GCGGCTTCAAGTCCGCGGCTTCCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4788  
Db GCGGCTTCAAGTCCGCGGCTTCCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5207  
Qy 5148 GCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5207  
Db GCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4728  
Qy 5208 ACCGCTGCTCCGCGGAGCACCGGCTTCCGAGTGAAGTCCGCGGAGCACCTTCCGAGTGAAGTGAAG 5267  
Db ACCGCTGCTCCGCGGAGCACCGGCTTCCGAGTGAAGTCCGCGGAGCACCTTCCGAGTGAAGTGAAG 4668  
Qy 5268 GGTTCCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5327  
Db GGTTCCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4608  
Qy 4667 GGTTCCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4608  
Db GGTTCCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5387  
Qy 5388 GGGTACGCGGAGCGGTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5387  
Db GGGTACGCGGAGCGGTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4548  
Qy 4607 GCGAGCGGAGCGGTCGATCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4548  
Db GCGAGCGGAGCGGTCGATCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4488  
Qy 4547 CCGGCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4488  
Db CCGGCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5507  
Qy 5448 CCGGCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4428  
Db CCGGCTGCTGCGGTCGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5507  
Qy 5508 GGGGAGAGATCCGCTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5567  
Db GGGGAGAGATCCGCTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4368  
Qy 4427 GGGGAGAGATCCGCTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4368  
Db GGGGAGAGATCCGCTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5627  
Qy 5568 GGTGAT 5627  
Db GGTGAT 4308  
Qy 4367 GGTGAT 4308  
Db GGTGAT 5628  
Qy 5628 TGGCTGCTGCGGCGGATTCACCGGCTGCGGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5627

Db	4307	TGCGGTGTCGCGGCGGAGATTCACCGGCTGGCGCTCGATATTCGGGACGCGGCGCGCGCT	4248
OY	5688	GACGAACTCACGTGTGGTGGTACGTCCGAGATCAGCTCAGCAGGGTGTCAATTGT	5747
Db	4247	GCAACAACTCGGCGCTGGCGGCGCAAGTCGAGATCAGCTCAGCAGGCTGTTCATGTGT	4188
OY	5748	GGAACGGGGGTGCGCGCCCGCAAGAGAAACGACGATCGGGTCTCGAAGAACCCCTGGG	5807
Db	4187	TGGCCGGGGTCCCGCGCCGATGAAAGATGACAGACCGGCGCTGGAAACAGCCCTCGG	4128
OY	5808	GCAGTCCGTCAGTACCCGCGCGACGAGCCCGGCTTCCACCTCGTCCGATATGACGCA	5867
Db	4127	GCAAGCGCGCGAGGACGTGCTGCGCGCGGCTCCCGCTCCACCGCGGTGGAGCGAGGCA	4068
OY	5868	GCAACGACCCGACAGGTCATATTTACACAGTTGTGCTAGCGAGCATCGCCCGCAGTGG	5927
Db	4067	GCAACGTCGCGAGGTCATATTTACACAGTTGTGCTAGTCAGCATCCCGCAGCTGCG	4008
OY	5928	GAGGGGTAGCCACTGCAATTCGGGGTCTCGGGGGGGTCCGCAAGACCTCGACCA	5987
Db	4007	GACCGGTACGCCACCGGAATTCGGGGTCTCGGGCAGGTGTTGAAACCTTCGACACCA	3948
OY	5988	TGTTGCGGTTGCGTTTGGCAGAAACAGAAACCTGTCTGACATTGAGGACGTCAACA	6047
Db	3947	TGTTCCGGTTCGGTTCGGCAGAAACACGCGCCCTGCTCGACCTGAGGACGTCAACA	3888
OY	6048	GGAATCCGGGGGGGGGGTCCCGTCGAAATCTCGATGAACTGGAACGCGGATCCGTTGT	6107
Db	3887	GGAATCCGGCTCGGGGGCGGTGCGCTTGAAGTACTGATGAAACCGGACCTTCAGCGCGGT	3828
OY	6108	GGAACCTCATGTATGTGCTGCGGGGTGCGCTCAGAGGTGCGGAGAGCTGACGCGATTGA	6167
Db	3827	GGAACCGCGGTATATGTGCTGCGGGTCCGCTCAGAGGTGCGGAGAGCTGACGCGATTGA	3768
OY	6168	TGTTGCCAAGTTTCGGGCTTGGCGCTTGGCACAGAGCGTGCATGACGCGCTCGACCT	6227
Db	3767	TGTTGCCCGGCTCGGCGCTTGGCGCTTGGCGCACAGTGCAGCACACCGTCGAACTCTTGA	3708
OY	6228	GCAATCAGACCGAGGAAACCGAATCTCGGGTTGGAAGATATGGATTGATCCAGTCCCGTC	6287
Db	3707	GCAATGAGCCGAGAGCCGAAATCTCGGGCTCAGATGATATGGGCTTGATTCAGTCCCGCC	3648
OY	6288	GCCATTCGAAATTTGATTCGGAAGTGCAGGCGCTCGATGAGAGAAAGCGCCCGAGTCTGT	6347
Db	3647	GCCAGCGGAATTTGATTCGGAAGGCGCTCGATGAGAGAAAGCGCCCGAGTCTGT	3588
OY	6348	GCGCCAGCCGACCGTCTCTCCGGTGGAAACGACACAGCGTTCCATATGTGCTGAAGGCACTC	6407
Db	3587	GCGGAGGCTTCCCGTCTCGGGCTTGGAAACGACACAGCGGTCCATGAGCGCTCGAACCGCACG	3528
OY	6408	GATGACCTCGAGCGGAATGCTCGGCGGGCGGGTGGGCAACCAATCGTGAACGTGCTCGG	6467
Db	3527	GCGGCACTTCAGACCGCGTCTCGGCGCGGTTGCGGGTCAAGCAATCGTGGGTTTCCCGCG	3468
OY	6468	TGGCGGTGGAGGTCCGCGCGTCCGAGTCAGAAACGTAATTGCTCGAATTGTGTGATTCG	6527
Db	3467	TGGCGGTGGAGGATTCGCGCGTCCGAGACAGAAACCGGTTGCGGTCTGGAGGATTCG	3408
OY	6528	GAGTCGCAATGACGTTGAACCCGATCCCCCATACGCTCTCCCGATATGTGTTGGGCGGTC	6587
Db	3407	GAAATGTCGCGGTGCGCGGTCCGATACCCATCGACTTCCCTTGGGAGTGTGTCACGCGCT	3348
OY	6588	CGTCCGGTACCGCCCGGACTGACATATTCGTGATCAAGCC-CGCGCAGTATAGGCTTC	6646
Db	3347	CCGCTTGTGTCGCGCGGGC-----GTTGTGATCAAAACCCGCACTCATCTTAAAGGCTC	3293
OY	6647	GCCCCGACCGGAGAAAGGCTC 6667	
Db	3292	GGTCAACGCGCGGAGAAAGGCTC 3272	

LOCUS	AX000487	8160 bp	DNA	linear	PAT 10-MAR-2000
DEFINITION	Sequence 6 from Patent WO9505283.				
ACCESSION	AX000487				
VERSION	AX000487.1	GI:7240896			
KEYWORDS					
SOURCE	unidentified				
ORGANISM	unidentified				
REFERENCE	unclassified sequences.				
AUTHORS	1 (bases 1 to 8160)				
TITLE	Mendez, C. and Salas, J. A.				
JOURNAL	BIOSYNTHESIS GENES AND TRANSFER OF 6-DEOXY-HEXOSES IN				
FEATUERS	SACCHAROPOLYSPORA ERYTHRAEA AND IN STREPTOMYCES ANTIBIOTICUS AND				
source	THEIR USE				
	Patent: WO 9905283-A 6 04-FEB-1999:				
	MENDEZ CARMEN (ES); SALAS JOSE A (ES)				
	Location/Qualifiers				
	1..8160				

```

mat peptide      242_
gene             2510..3223
                /gene="ERICVT"
CDS             2510..3223
                /gene="ERICVT"
                /function="IMPLIQUE DANS LA BIOSYNTHESE DE LA DESOSAMINE"
                /codon_start=1
                /protein_id="CAB7094.1"
                /db_xref="GI:7240897"
                /translation="MYEGGFAELVDFPYRGGKDYAAEAQAQVRLVDRILPSASLLD
VACGGTIRRRPADLFDVYGLIELSAMIEVARPOLGIPVLQDMDFPALDRFDFAV
TCMFSSIGHMDGALDQALASPARHLAPGVVVEEPMWPFEDLDQGVVAGDVVRDSD
LTISVSSVAGGATGMEITHVAVADAVNGPRHHVEHYELTLPERQYKRAFTAGCA
VOYLEGSGSGGLPLVQVVG"
gene            6080..7549
                /gene="ERICV"
CDS            6080..7549

```

```

//function="IMPLIQUE DANS LA BIOSYNTHESE DE LA DESOSAMINE"
//codon_start=-1
//protein_id="CAB7095.1"
//db_xref="GI:7240896"
//translation="MNTTRTATQAGVADAAAPDVDRRAVRAVLSSEVTSYAGAGG
DAVOAARLADLAAHYGAPHFPTBOTRRALGIDRAEPHLIDLPRIPIGARVEA
PARKYSNTKPIIDAGALDAPYRRKPARPYSGULRGPCIMFRCHCYRATVAYEA
ASVPAGNETTAAIIVDPIDNPAMWSGGLVTPNGEIVSHAAGRFDTIVYTIN
AFAPLTQTLNRPGWGLWGLAIRSLGLVNLDEYTTGKGAERYKQNGGLRMA
ERAPAPRLGNNHILPGRADRLTDLPVDFIAELNSESPPORDLVTVRDEYSGDDGL
SDSERHLEEGIVRFPDYAABRRPGMHIDUGVALBESLRGVADDELARIPIETMRPAH
PQVAVIDLIDGDIYLVREAGFPELEGATRIIACGVTPSTLSREVENFVLENGVOPR
PGGEVYIDGDSVTRALNQLEBIDADGWEHDHNGFRRGR"

```

[illegible]

QY 2508 GCGAGTGAATGCTGCGGCGCCGACGTAACGAAGTCGTGAACGACCGAGTGCAGGCTCGGAT 2567  
DB 7427 GGGGCTGACGCGCTCGTTCTTCCAGCAAGAAATTCTCCACACCTCCGCGAGGCTGTGTCG 7368  
QY 2568 CCGGGGGTGAACCGGCGCGGACGTAAGCGGGTGGCCCCCGCCAGCCCGGGGAAACCGGCT 2627  
DB 7367 ACGGGGGTACCCGCGCGGACGTAATGAGCGGGTGGCCCTCCAGCTCCGGGAAAGCCCTT 7308  
QY 2628 CCGGTAACAGGTACGCTCTCCAGACGAGTGCATCTGCAACGCGACCTGCGGATGGGCGG 2687  
DB 7307 CGGGTACAGGTACGCTCCGCGAGCAGTGCATCTGCAACGCGACCTTGGGGGGTGGCGG 7248  
QY 2688 TCGGACGATCTGTTCCGCGCGAATCCGCAACAGCTGGCGTGCACACCCCGACGACGC 2747  
DB 7247 TGGGACGATCTGCTCCGCGCGGAAATGCGACAGACTCGCGGTCCACACCCCGCGCAGGC 7188  
QY 2748 TCTCAACGCGTAACCCAGAGTGGTCTGCAATGCCCGGGTCCGCTCGCGCGGTAATGCA 2807  
DB 7187 TCTCAGGGGCTACCCAGAGTGCATGTCATGCGCCGGGTCCGCTCGCGCGGTAATGCA 7128  
QY 2808 CGAACCGGGGAGGCGCTCCCGGCACTTCGTCGGTTCGCTCCGGAACAACCTGCGGT 2867  
DB 7127 CGAACCGGACCGGCGCTCCGCGAGCTCTGTCGCTCGAGTCCGACAGCCGCGCTGCT 7068  
QY 2868 CCGGACCGCTGTAGTCTCTCGCGACGCTGACGAAGTGCAGCGCGCGCTGCGGCTGACCT 2927  
DB 7067 CGCGCGGCTGTAGTCTCTCGCGACCGCTGACGAAGTCCAGCGCGCGCTGTGGCGGCTGACCT 7008  
QY 2928 CGTTGAGTTCGCGCAATGAATGCACAGGTGCATGAGCGGGTGGCCCGCCCGCAGGA 2987  
DB 7007 CGTTGAGCTCGGCAATGAATGCACAGGTGCATGAGCGGGTGGCCCGCTCCCGCAGGA 6948  
QY 2988 TGATGTGGTTGAGGCGGACCGGACCGGTGCGCGCTTCGGCGCGCATCCGCAAGAAAT 3047  
DB 6947 TGATGTGGTTGAGGCGGACCGGACCGGTGCGCGCTTCGGCGCGCATCCGCAAGAAAT 6888  
QY 3048 CCGGAGGTTCTTCTGACCGCTTCGAAGCGCGCACGCTTCGCGGTGAGCTGTGATCT 3107  
DB 6887 CTTGAGGTTCTTCTGACCGCTTCGAAGCGCGCGCGCTTCGCGGTGAGCTGTGATCT 6828  
QY 3108 CGTGGTGTTCAGCCCGTAACAGGAGGTACGGAACCGCGCGCAAGTCCCAAGCGCGGCT 3167  
DB 6827 CGTGGTGTTCAGCCCGTAACAGGAGGTACGGAATCGCGCCCAAGCTCCCAAGCGCGGCT 6768  
QY 3168 GCGCGCGAGGCTGTTCGTAAGGAGGCGGAAGCGTTGGTGTGAGACGCTGCAGGGCGAAGC 3227  
DB 6767 GCGCGGTAGGCTGTTCGTAAGGAGGCGGAAGCGTTGGTGTGAGACGCTGCAGGCGAAGC 6708  
QY 3228 CCGCGCGGCGCGGTCGACACCAACGCGCCGATACCGGGTGGTCAACCGGTTCCAGGC 3287  
DB 6707 CGCGCGCGCGCGGTCGACACCAAGCTGCCGAACCGGGGTGGTCAACCGGCTCCAGGC 6648  
QY 3288 CGCGGAGAGGTACATGCGCGGTTCGCGGTTCTTCGACGCGACCTTCGTGATGACCGAGTGA 3347  
DB 6647 CGCGCGGATGTACATGCGCTTCGCGGTTGTCCGTGAGGACCTTCGTGATGACCGAGCA 6588  
QY 3348 GGAATCTGTTGCGCGGACACACCACTCCGCGTGTACTTTGGACACTGTACCCGTAGGC 3407  
DB 6587 GCGTCTGTTGCGCGGACACCACTCCGCGTGTACTTTGGACACTGTACCCGTAGGC 6528  
QY 3408 AGAAGTACAGCGGAACATGACGTCGCGGCGCGGGTACAGCCCGACGCTGTACGGAAG 3467  
DB 6527 AGAAGTACAGCGGAACATGACGTCGCGGCGCGGGTACAGCCCGACGCTGTACGGAAG 6468  
QY 3468 CGGGTCCCGCGTAACGCGGCTTCGAACGCGCGCGGCGGCTTCAGCGCGCTGATCGGT 3527  
DB 6467 CAGGCTTCGCGTAACGCGGCGTTCGAAGCGCTTCGCGCGCTTCGAGCGCTGATCGGT 6408  
QY 3528 TGGTCAATATTCGCGCGCGCGCGCTTCGACCGCGGTGCGCGCTTCGCGGAGTGCAT 3587  
DB 6407 TGGACCAATATTCGCGCGCGCGCGCTTCGACCGCGGTGCGCGCTTCGCGGAGTGCAT 6348  
QY 3588 CGAACAGGTGACGACTGCGCGAACCGCGCGCGGTTCAGGTCGAACCGACGCGCATCT 3647

DB 6347 CGAACAGGTGACGACTGCGCAACTCCGCGGCTTCAGAGCGGACCGGACGCGCT 6288  
QY 3648 GCTCCAAACGGGGTAACCGCGAGCGCGCGTAATGCGGCGGTAATTCGCGAGCGGACCG 3707  
DB 6287 GCTCCAAACGGGGTAACCGGAGTGGCGCCCGTAATGCGGCGGTAATTCGCGAGCGGCGG 6228  
QY 3708 CTTGTGTTGCGCGGTAATGCGGACCGGTGAGCGGTAATTCGCGGCTGTAAGC 3767  
DB 6227 CTTGACAGTGGCGGTAATTCGCGGACCGGCGCGGTAATTCGCGGAGTCAAGC 6168  
QY 3768 CCGGACACAGCGCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 3827  
DB 6167 CCGGACACAGCGCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 6108  
QY 3828 CCGGCGGCGGTAATGCGGTAATTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 3887  
DB 6107 CCGGCGGCGGTAATGCGGTAATTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 6048  
QY 3888 GAGATACAGAACCGGTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 3947  
DB 6047 GGTGTACAGAACCGGTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5988  
QY 3948 GTGGTGGCGAGTGGATGACGTCGACACCCCGCGGATGTCCTCACTGACACCGGAGG 4007  
DB 5987 GTGGCGGCGAGCGGATGACGTCGACACCCCGCGGATGTCCTCACTGACACCGGAGG 5928  
QY 4008 CCGGTGCGGAGGCGGACACCCGTTGCGGAGCGCACTCGGCTGTGTCAGCGGAGCGGT 4067  
DB 5927 CCGGTGCGGAGGCGGACACCCGTTGCGGAGCGCACTCGGCTGTGTCAGCGGAGCGGT 5868  
QY 4068 GGTCCGTGCGGTAATGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4127  
DB 5867 GGTCCGTGCGGTAATGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5808  
QY 4128 ACTTCTTCGCGGCGGAGATGCGCTTCGACCGCGCTTCGAGTGCATGCGGCGGCGGCTTCGCGG 4187  
DB 5807 ACTTCTTCGCGGCGGAGATGCGCTTCGACCGCGCTTCGAGTGCATGCGGCGGCGGCTTCGCGG 5748  
QY 4188 TCCACAGAGTAATGACGTAATGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4247  
DB 5747 TCCACAGAGTAATGACGTAATGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5688  
QY 4248 GTGACCGCGGTAATGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4307  
DB 5687 GATATGCGCGGTAATGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5628  
QY 4308 TCCGTGCTTCGCGGAAACGCGTTCGAGGAGTGAAGCCCATGCGCGGCGGACCTTCGCTC 4367  
DB 5627 ACCGTGCTTCGCGGAAACGCGTTCGAGGAGTGAAGCCCATGCGCGGCGGACCTTCGCTC 5568  
QY 4368 ATCTTGGCTTCGCGGAGTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4427  
DB 5567 ATCTTGGCTTCGCGGAGTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5508  
QY 4428 ATGCGCGGATCGGTTGCGGACGAGCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4487  
DB 5507 ATGCGCGGATCGGTTGCGGACGAGCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5448  
QY 4488 GCGGTGACCGGCTTCGCGGAGTGAAGCTGAACCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4547  
DB 5447 GAGGTGACCGGCTTCGCGGAGTGAAGCTGAACCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5388  
QY 4548 GGAAGTCAACCGGCTTCGCGGAGTGAAGCTGAACCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4607  
DB 5387 GCGCGCGCGCGGCGGTAACCGGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5328  
QY 4608 TGGTGGCGGAGTTCGCGGAGGCTTCGACAGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 4667  
DB 5327 TGGTGGCGGAGTTCGCGGAGGCTTCGACAGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGGCTTCGCGG 5268  
QY 4668 CCGACGATCGCGCGGCTTCGCGGAGTGAAGCGCGGCGGAGCTGCGTTCGCGGCTTCGCGGCTTCGCGG 4727

Dp	5267	CCGATGATCGCCCGCTCCGCGGTGTCACAGCCAGCGACGTGCTCGGGTCCAGCAGG	5208
Oy	4728	CCGGTCGCCGGTTCGACGTTCGAGAACACCGAGTGAATCCGAGCAGACTCGCCGTCGG	4787
Dp	5207	CCGGCTCGGGGGTCCACGTTCGAGAACACCGGGTTCAGCCCCAGCAGACTCGCCGTCGG	5148
Oy	4788	GGGGTGGCCCGGAAGTCAATTGACGGCATGATCACTTCAACCGGTGAAGTTCACCCGACCG	4847
Dp	5147	GGGGTGGCCCGGAAGTCAATGACAGCTGACAGCTGCGCGACACCTGCTCGCGGC	5088
Oy	4848	AGCACCAATTCCAGGGCCGACGGTGGCGTTGACAGGTGGGAGTGGCGATGCGACCCGAC	4907
Dp	5087	AGCACCAATTGACGGCCGACCGTGGCGTTGAGGTGGGACGACAGTGGCCGACACCGCC	5028
Oy	4908	AGGTCCGGGACCCCGGCTCCGAACTCCCGACACAGGGGCCCGCCGTTGGTAGGCAATTG	4967
Dp	5027	AGGTCCGGGACCCCGGCTCCGAACTCCCGGACACAGTGGTCCGCGTTGGTCAAGCAATTG	4968
Oy	4968	TTGTTCAGGGCCCACTTCAGGGCGGGCCAGGAACGGCTGCGCGGTGCGCATGTCGCGCG	5027
Dp	4967	TTGTTCAGCGCCCACTCCAGGGCGGGCCAGGAACCGCTCCCGGTCCCGACGTCGCGCTG	4908
Oy	5028	CCCAAGTCAGGGGGTGGAGGAACGCTCCGGGACCGCGGAAGTCCGACATCCGATCCGAT	5087
Dp	4907	CCCAAGTAGAGGGGTGGAGGAATGCTCCGGGGCCGCGGAAGTCCGACGATTCGATCAGC	4848
Oy	5088	ACGCGCTTCAACGCGCTCCCGGTTGTGACGCGCGGACGCGCAGGCGGACGAGGCTGCG	5147
Dp	4847	GGCGGTTTCATCCGCGCGTCTTGATGATGACGCGCGAGGCGCACGCGATGATGCTGCGCA	4788
Oy	5148	GCTGATTTTGAAGTATGCTGTGTCGCGGACAGTTCGATCTGCGTGGTCCGAGGTCATCC	5207
Dp	4787	GCTGATTTTGAAGTATGCTGTGTCGCGGACAGTATGATCTGCGTGGTCCGAGGTCATCC	4728
Oy	5208	ACCGGTGGTCCGGGGCCACCGCGCTCCCTGAAAGTCCGCGGGGACCTTCGACGAGCATGAC	5267
Dp	4727	ACCGGTGGTCCGGGAGCGGCACTGGCGTGAAGTCCGCGGGGACCTTCGATCGAGCATGAC	4668
Oy	5268	GGTTCCTGTTGCGGTAGAACCGCCCACTCTCGAGATGCGACGCGTCCGTAAGGAGAACT	5327
Dp	4667	GGTTCCTGTTGCGGTAGAACCGCGCGCTCTCGAGATGCGACGCTCGTAAGGAGAACT	4608
Oy	5328	GGGTACCGGGGACGTCAGACAGTATGCGAGTAAAGTCGGCGGGTGTCTTCGCGGACGT	5387
Dp	4607	GCGAGCGCGGACGTTCAACAGTATGTCACATATGGCCGCGTCCGGAACCTCCCGGGCGGT	4548
Oy	5388	CGGTGATGTTGTCCGGCTGACAGTGCACCGTCCGCGGCAACTCGGCGACGTCGTGGCAC	5447
Dp	4547	CGGGGTGATGTTGCGGGCTGGGAGTGCACCGTAAAGCGGCACTTCGCGACGTCATCCGCG	4488
Oy	5448	CCGCTCCGATGCGAGGTCAGAGCGGTCAGAGGGTGGCGGTGATCTCTGTAACAGGA	5507
Dp	4487	CCGCTCCGATGCGAGCTGACAGCGGTCAGACGTCGTGCCGATGTCTTGAACAGGA	4428
Oy	5508	GGGCGAGCAGTCGCTGTGTTGGCGGGGGAGAGCAGCGGTTGCATCAAGAACCTCACTCGC	5567
Dp	4427	GGGCGAGCAGGCGCTGTTGTTGGCGGGGAGAGCAGCGGCTGCATCAAGATGACCTCGC	4368
Oy	5568	GGTACATGTGTGTCACGGACGACCGCCGAAGATGCTGAAGTCTTGTCTCTCGTGTGCA	5627
Dp	4367	GGTGTGTGTGTGCAACCGTGAACCGCCGAAGATGCTGAAGTCTTGTCTCTCGTGTGCA	4308
Oy	5628	TGCGGTGATGCGCGGATCCACCGCGTGGCGGTGATGTGGGCGGAAGGCGCGGGTGGCTT	5687
Dp	4307	TGCGGTGATGCGCGGCGATCCACCGCTGGCTGATGTGGGCGAGCGCGGGCGCGGCT	4248
Oy	5688	GGAAGAACTCACTGTGTGTGTGATGCTCCAGATCCAGCTCAGCAGGGTGTCAATTGT	5747
Dp	4247	GCAAGAACTCGCGCTGGCGGACGTCGAGATCCAGCTCAGCAGGCTGTGATGATGTT	4188
Oy	5748	GGAACGGGGTGGCCGCCCGACAGGAAGACGACAGATCGGGTTCGAAAGACCCCTCGG	5807
Dp	4187	TGGCGGGGGTGGCCGCCCGATGAACAGATGTCAGACAGCGGGCTTGAAAGAGCCCTCGG	4128

QY	5808	GCAAGTCGCTCCAGTAACCCGCGCGAAGCGCCCGGCTTTCACCTCTGTGGTATGCAAGGCCA	5867
Db	4127	GCAAGCGCGCGACAGCAACGTGTCGCGCCCGGTCTCCGCTCCACCGCGGTGGGACGCAAGGCCA	4068
QY	5868	GCAACGGAACCGACAGGTCATATTTCACAACGTTGTGTATGCGGAGGATGCGCCCGACGTGGG	5927
Db	4067	GCAACGCTGCGAGGTCCATATTTTCCACAGTTGTGTATGTCAGATGACGCGCCGACGCTGCG	4008
QY	5928	CGAAGGTGAGCAACCTGCAAGTTTCGGGTGTCTCGGCGGGTCTTCGAAAGCCTCGACCA	5987
Db	4007	CGACGCTCAGCAACCGAAAGTTTCGGGTGTCTCGGCAAGTGTGTGACCACTCGACGACCA	3948
QY	5988	TGTTTCGGTTTGGCTTTGGCAGAAACCAAGAACCTCTGCTGGAATGAGAGCTGCACGA	6047
Db	3947	TGTTTCGGTTTGGCTTTGGCAGAAACCAACGCGCTCTGCTCGAATGAGAGCACTGCACGA	3888
QY	6048	GGATCCGCTGGGGGGCGGTCCTCCGTCGAAGTATCTGAAATCTGAAACCTGAAACCGCGGATCCGTTGT	6107
Db	3887	GGATCCGCTCGGGCGCGCTGCTGCTTGAAGTATCTGAAATCTGAAACCTGGAACCTTCGAGCGCGGT	3828
QY	6108	GGAACCTCATATAGTTGCTGTGCGGGTGTGACCTTGACAGAGGTTCGAGAGAGCTTCGACGCGCTTGA	6167
Db	3827	GGAACCGCGGTATAGTTGCTGTGCGGGTGTGACCTTCGAGAGAGCTTCGAGAGAGCTTCGACGCGCTTGA	3768
QY	6168	TGTTTCGACAGTTTCGGGCTTGTGCTTGACCAACAGCGGTGACGTAACGCGCTGAACTTCCGCA	6227
Db	3767	TGTTTCGCGGCTCGGACTTGAGCTTGAGCCCAAGCAATGTCAGACACACGCTGCAACTCTTTGA	3708
QY	6228	CGATCAGAACCGAAGAACCCCATCTTCGGGTTTGAACGATGATGGTTGGAATCCAGTCCCGTC	6287
Db	3707	CGATGAGGCCACAGAAACCCCATCTTCGGGCTTGACAGATGATGGGCTGATCCAGTCCCGTC	3648
QY	6288	GCCATCCGAAATTTGATGTCGAGCGTGCAGAGGCGCTCGATGAGAAAGAACGCGCCGGAAGTCGT	6347
Db	3647	GCCAGCCGAAATTTGTCGCGACGTCGAGGCGCTTCGATGAGAAAGAACGCGCCGGAAGTCGT	3588
QY	6348	GCGCCAGCCGACCGTCCCTCGGGGTGGAACGACCAAGCTTTCATATGTTCTGAAGGGCACTC	6407
Db	3587	GCGGAGAGCTGTCGCTCTCGGGGTGGAACGACCAAGCGGTTCATATGCTGAACGGCACTC	3528
QY	6408	GGTGACCTCGAGCGCATGCTTCGGCGGCGCGGTGGGCGCAACAGTCGTGGAACGTCTGTGCG	6467
Db	3527	GCGCCACTCTCAGAGCGCGTCTCGGCGCGCTTTCGGGGTCAACAGTCGTGAGTTTTCGCGCG	3468
QY	6468	TGGGCGGTGGAGGTTCGCGCGGTGCGCATCAGAGAAACGTATTTGCGCATTTGTGTGATTTCCG	6527
Db	3467	TGGGCGGTGGAGATTCGCGCGGTGCGCATCAGAGAAACGCTTTGCGGTCTGGAAGATTTCCG	3408
QY	6528	GAGTCGATGATCAGTTGATCCGATATCCCATATGAGCGCTCTCCGCTGATGATGTCGAGGCGGTC	6587
Db	3407	GATTCGTCGCGGTGCGCGGTCCGATCAACCATTCGACTCTCCCTGCGGTGTGTGTGACGGCT	3348
QY	6588	CGTCCGCTAACGCGCCGGAATGACATTTCTGTGATCAAGACC-CGCCGATGTAGGGCTCC	6646
Db	3347	CGCGCTTGTGCGCGCGCGGCG-----GTTGTGATCAAGAACCCGCACTCACTGTAAAGGTCC	3293
QY	6647	GCCCGCGACGAGGAAGTCC	6667
Db	3292	GGTACGCGCGAGGAAGGCGC	3272

RESULT 4	AR092722/c	8051 bp	DNA	linear	PAT 08-SEP-2000
LOCUS	AR092722				
DEFINITION	Sequence	2 from patent US 5998154.			
ACCESSION	AR092722				
VERSION	AR092722.1	GI:10019474			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 8051)				





Db 5524 GATATGCGCGCAGCTGCGAGAGCTCGAGATGAGCGGTGATGAGCGGATTTGTC 5465  
Qy 4308 TCCCTGCTCGCGAGAACGCGTGAAGAGAGTGAAGCCCCATGACCGCGGCGACTGCTC 4367  
Db 5464 ACCCTGATCTCGCGAGAGCGCTGAGAGAGTGAAGCCATCCCGCGGCGACTGCTC 5405  
Qy 4368 ATCTGCGGTTGCTGAGCATCTGCGTGAACACCTTGTCCGCGCGAGATGCGAAGTTGTG 4427  
Db 5404 ATCTTGCCTGTTGATGCGAGATGAGTCAACAGCTTGTCCGCTGCGATCCGAAAGTTGTG 5345  
Qy 4428 ATGAGCCCGAATCCGTTGCGGCGAGAGAGTGAACACCTGCGCTGCACTCGATCCGCCAC 4487  
Db 5344 ATGAGCCCGAATGCGGCTGCGGCGAGAGAGCTGCGCTGCGTGAAGATGAGCGGCTGCAAC 5285  
Qy 4488 GCGGTGACCGGCTGAGTGGGTGAGAGAGTGAACACCTGCGCTGCACTCGATCCGCCAC 4547  
Db 5284 GAGGTGACCGGCTTCTGAGGTGAGAGAGTGAACACCTGCGCTGCGTGAAGAGCGGCGAC 5225  
Qy 4548 GGAAGTCAACCCGTCGTCAGCGCGAGAGCGTGGCGGCGCTGAGAGAGAGCTTGAACCTG 4607  
Db 5224 GCGCGGCGCGCGCGGCTGAGCGCGAGAGCGTGGCGGCGCTGAGAGAGAGTGAACCTG 5165  
Qy 4608 TGGTCGCGCGAGATCTTTCGCGAGCGCTGCAAGATGCTGTCGCGCGCGCAAGTGAAC 4667  
Db 5164 TGGTCGCGCGAGATCTTTCGAGCGCTGCAAGAGCGGCGCTGCGCGCGCAAGTGAAC 5105  
Qy 4668 CCGAGCATGCGCGCGCTGTCGCGGAGTGAAGAGCGCGCGAGCGTGGCTCGAGTGAAC 4727  
Db 5104 CCGATATGCGCGCGCTTCCGCGGTGAGACAGAGCGAGCGAGCTGTCGCGGTGAGACAG 5045  
Qy 4728 CCGGTGCGCGGCTGAGAGTGAACAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4787  
Db 5044 CCGGTTCGCGGCTGAGAGTGAACAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4985  
Qy 4788 GCGGTGCGCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4847  
Db 4984 GCGGTGCGCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4925  
Qy 4848 AGCATCAAGTTCAGAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4907  
Db 4924 AGCATCAAGTTCAGAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4865  
Qy 4908 AGGTGCGCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4967  
Db 4864 AGGTGCGCGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4805  
Qy 4968 TTTTCAAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5027  
Db 4804 TTTTCAAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4745  
Qy 5028 CCGCATGAGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5087  
Db 4744 CCGCATGAGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4685  
Qy 5088 AGCGGCTTGAAGCGGCTGCGGCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5147  
Db 4684 AGCGGCTTGAAGCGGCTGCGGCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4625  
Qy 5148 GCTGAGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5207  
Db 4624 GCTGAGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4565  
Qy 5208 ACCGCTGCTGCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5267  
Db 4564 ACCGCTGCTGCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4505  
Qy 5268 GGTTCCTGTTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5327  
Db 4504 GGTTCCTGTTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4445  
Qy 5328 GGGATGCGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5387

Db 4444 GCGAGCGCGAGAGTTCACACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4385  
Qy 5388 CGGTGAGTGAAGTTCGCGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5447  
Db 4384 CGGTGAGTGAAGTTCGCGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4325  
Qy 5448 CCGCTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5507  
Db 4324 CCGCTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4265  
Qy 5508 GCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5567  
Db 4264 GCGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4205  
Qy 5568 GGTGATGCTGCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5627  
Db 4204 GGTGATGCTGCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4145  
Qy 5628 TGGCGTCTGCGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5687  
Db 4144 TGGCGTCTGCGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4085  
Qy 5688 GGAAGAACTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5747  
Db 4084 GGAAGAACTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4025  
Qy 5748 GGAAGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5807  
Db 4024 TGGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3965  
Qy 5808 GGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5867  
Db 3964 GGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3905  
Qy 5868 GGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5927  
Db 3904 GGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3845  
Qy 5928 GGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5987  
Db 3844 GGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3785  
Qy 5988 TGTGCGGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6047  
Db 3784 TGTGCGGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3725  
Qy 6048 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6107  
Db 3724 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3665  
Qy 6108 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6167  
Db 3664 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3605  
Qy 6168 TGTGCGGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6227  
Db 3604 TGTGCGGTTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3545  
Qy 6228 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6287  
Db 3544 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3485  
Qy 6288 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6347  
Db 3484 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3425  
Qy 6348 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6407  
Db 3424 GGAATCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3365  
Qy 6408 GGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6467  
Db 3364 GGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3305









Db	4483	GGTGGCTGCTGGCCACCGTGAACCGGAAAGACTGCAGAACTTCTTCTCTGCTGCA	4422
OY	5628	TGCGGTGCTGCGGCCGGAATCCACCCGCTGCGGTGATGTGCGCCAAACGCGCGGGTGCCTT	5687
Db	4423	TGCGGTGCTGCGCGGGAATCCACCCGCTGCGGTGATGTGCGGGCAAGCGGCGCGCT	4366
OY	5688	GGACGAACTCACTGTGTGGGTAGTGTCCAGATTCACACTCAGCAGGGTGTTCATGTTGT	5747
Db	4353	GACCAAACTCGGCTGTGGCGGGCAAGTGCAGAGATCAAGCTCAGCAGGGCTGTTCATGTTGT	4304
OY	5748	GAACCGGAGGTACCCGCCCCGACGAAAGAACGACGAGTCCGGGTTCGAAAGAACCCCTGGG	5807
Db	4303	TGCGCGGGGTGTCCCGCGCCGATGAACAGTGTACAGACCGGGCTTGGAACAGGCCCTTGG	4244
OY	5808	GCAAGTCCGTCCAGTACCCGGCCGACAGGCCCCGGCTTCCACCTGTGTGCGTATGCAAGCCA	5867
Db	4243	GCAAGCGCGCGCAGACAGTGTGTGCGGCCGGTCCCGCTCCACCGCGGTGTGGAGCGACGCA	4184
OY	5868	GCAAGGACCGCAGAGTCCATGTTTACCAACAGTTGTGCTAGCGAGCATGCGCCGCAATGGG	5927
Db	4183	GCAAGGTGTCCGAGGTCCATGTTTCAACAGTTGTGCTGTGTGACAGCATCCCGCAGTGTGG	4124
OY	5928	CGAGGGTGAAGCACTGTGAGTTCCGGGTGTGCGGGCGGGTCCCTCGAAGACCTTCGACCA	5987
Db	4123	CGACGGTTCAGCCACCGAAGTTCCGGGTGTGCGGGAGGTGTGTCAACCTTCGACGACCA	4064
OY	5988	TGTTGCGGTTGCGTTTGGCAGAAACAGAAACCTGTGTGCGACTGTGAGAGACGTTCACGA	6047
Db	4063	TGTTTCCGGTTGCGTTTGGCAGAAACACCGGCCCTGTCCGACTGTGAGACAGTTCACGA	4004
OY	6048	GGATTCGGTGGGGGCGGGTCCCGTCGAAAGTACTGATGAAGTGTGACGCGGATCCGTTGT	6107
Db	4003	GGATTCGGGCTCGGGCGCGCTGCGCTGTGAAGTCTGATGAACCGGACCTTCGACGCGGGT	3944
OY	6108	GGACCCCTCATGTAAGTGTGCGGGGTGCGTCCAGAGTCCGCGAAGCTGTGAACCGCGTTTGA	6167
Db	3943	GGAGCGCGGTGTAGTGTGTGCGGGGTGCGCTCCAGAGTCCGCGAAGCTGTGAACCGCGTTTGA	3884
OY	6168	TGTTGCGCAGTGTCCGAGCTTTGGCCTTGGCCAGACAGAGGTGTGATGCGCGGTCCGACCTCCGCA	6227
Db	3883	TGTTGCGCGGCTCGGGCTTGGGCTGTGGCCTGTGCGCAGCAGTGTGACACACCGTTCGAACTCTTGA	3824
OY	6228	CGATTCAGACCGAGGAACCCGATCTCCGGTTTGAACGATGATGGGTTGATTCAGTCCGTC	6287
Db	3823	CGATGAGGCGAGGAAGCCGATCTCCGGCTGTGACGATGATGGGCTGATTCAGTCCGTC	3764
OY	6288	GCCATTCGGAATTTGATCCGACAGTGTGCAAGGCCCTTGAATGAGAAAGCGCCCGAGTGT	6347
Db	3763	GCCAGCCGAATTTGTCCGACAGTGTGCAAGGCCCTTGAATGAGAAAGCGCCCGAGTGT	3704
OY	6348	GCGCCAGCCGACCGTCCCTCCGGGGTGAACGACAGCGGTTCCCATGTTGTCTGAAGGCACTC	6407
Db	3703	GCGCGAGCTGTGCGCTCTCCGGGCTGGAACGACAGCGGTTCCATGAGCTGGAACGCAACG	3644
OY	6408	GGTTCACCTCCAGCCGATGTCCGCGCGCGCGCTGTGGGCGCAACCAAGTGTGTGACGTCCGTCCG	6467
Db	3643	GCGCCACCTCCAGCCGCGGTGTCCGCGCGCGCTGTGGGCGCAACCAAGTGTGTGACGTCCGTCCG	3584
OY	6468	TGCGCGGTGGAGGTCCCGCGTGTCCGAGTCAGGAAACGTAATTTGCCGATTTGTGTGATTTCCG	6527
Db	3583	TGCGCGGTGGAGATTTCCCGCTGTCCGAGCAGGAAACCGTTTGCAGGATTTCCG	3524
OY	6528	GAGTCGCAATGACGTTGACCCGATCCCCCATATGCGCTCTCCCGATGTGTGTGGCGGCTC	6587
Db	3523	GAAATGTCTCGGTGCGCGGTTCAGATCCCATGCACTCTCTGTGGGTGTGTGTGACGAGCT	3464
OY	6588	CGTGCAGTATCCGCCCGGACCTGACATTTGATGTGATCAAGACC-CGCGCAGTGTAGGGCTCC	6646
Db	3463	CGCGTTGCTGTGCGCGGGC-----GTTGTGATCAAGCCGACCTCACTGTATAGGGTCC	3409
OY	6647	GCCCCGACCGGAGAAAGTCC	6667

DB	3408	GGTACGCGCGGGAAGGCC	3358
RESULT 6			
LOCUS	AF263245	47981 bp	DNA linear BCT 17-SEP-2000
DEFINITION	Micromonospora megalomicea subsp. nigra megalomicin biosynthetic gene cluster, partial sequence.		
ACCESSION	AF263245		
VERSION	AF263245.1	GI:10179840	
KEYWORDS			
SOURCE			
ORGANISM	Micromonospora megalomicea subsp. nigra		
REFERENCE	Micromonospora megalomicea subsp. nigra		
AUTHORS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
TITLE	Micromonosporineae; Micromonosporaceae; Micromonospora.		
	1 (bases 1 to 47981)		
	Volchegursky, Y., Hu, Z., Ketz, L. and McDaniel, R.		
	Biosynthesis of the anti-parasitic agent megalomicin:		
	transformation of erythromycin to megalomicin in Saccharopolyspora		
	erythroa		
JOURNAL	Mol. Microbiol. 37 (4), 752-762 (2000)		
PIRMEED	10972798		
REFERENCES	2 (bases 1 to 47981)		
AUTHORS	McDaniel, R. and Volchegursky, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-MAY-2000) Kossan Biosciences, Inc., 3828 Bay Center		
	Place, Hayward, CA 94545, USA		
FEATURES	Location/Qualifiers		
SOURCE	1..47981		
	/organism="Micromonospora megalomicea subsp. nigra"		
	/mol_type="genomic DNA"		
	/strain="NRRL3275"		
	/sub_species="nigra"		
	/db_xref="taxon:136926"		
	complement (<1..144)		
	/gene="meg1"		
	complement (<1..144)		
	/gene="megT"		
	/codon_start=1		
	/trnal_table=11		
	/product="TDP-4-keto-6-deoxyhexose-2,3-dehydratase"		
	/protein_id="AAG13906.1"		
	/db_xref="GI:10179841"		
	/translation="MGDRVVGHAHPSTQSAIFLFRHGSPPTATDDVDHMLARAAE		
	HRLS"		
	928..2061		
	/gene="megDVI"		
	928..2061		
	/gene="megDVI"		
	/codon_start=1		
	/trnal_table=11		
	/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"		
	/protein_id="AAG13907.1"		
	/db_xref="GI:10179842"		
	/translation="MAVGDRRLRGRELOMARGLYMGFGANGDLYSMLSGRDDPPTW		
	YELRLAAGRPVPSRAGTVVGGHRTAAEYIADPGFHGPPDARMVQVHCPAASPA		
	GPREPVRATEDASVTADADYCOORCAFLYTELISGRPLDYNAPARVPVLAICTGTA		
	LKVVDRLARSATRVCLDAVYSPQCLAVTQCLTALDEIYVAGRRALVAGV		
	ASLAAVTGNAVLAVTELPELRLADDEPTARVTEVSRTSPGVHLERTTASDR		
	VGVVDVPGEVTVVVAANRDPEVFTDPDPVDGDAEILISRGSPTDLDLV		
	ATLATAALPAARVLPRLSRSGPVIIRRRSPVARGLSRCEVEL"		
	2072..3382		
	/gene="megDI"		
	2072..3382		
	/gene="megDI"		
	/codon_start=1		
	/trnal_table=11		
	/product="TDP-megosamine glycosyltransferase"		
	/protein_id="AAG13908.1"		
	/db_xref="GI:10179843"		
	/translation="NRVVPSSNAVNSHLGVLPLASFAFOAHEVAVVSPALTDVTV		
	GALTVLPVGGDVELVEWHAHAQDILVEYMRITLDVNDQSHITMSMDLLGMOTTFPT		
	PFLAMPSPDILIDQWVEFCRSWRDWTWSELTFLPAADIAARVTCGTPAHMIMGPDVATR		

gene  
CDS  
3462. .4634  
/gene="megy"  
/codon\_start=1  
/transl\_table=11  
/product="mycarose O-acyltransferase"  
/protein\_id="AAC13909.1"  
/db\_xref="GI:10179844"  
/translacion="MWTSTNLDTTARPAALNSLTGMRFVAFLVEFTVLSRLPNSVY  
YADGIDAFKQTVGVSFPFIISGPVLITMSABASDVSMSFRRVCKLPNNHLYTAP  
AAVVLPLVTQAVSGEALIPMLLIHMFPLETISFGINPYSGLACBAFFYLCFPIF  
LFMIISGIRPERLMAAIVPAIIVAVVADLLPSSPPLIPLEYSAIQDWFLYTP  
ATRIEFLIGIILRIITGTWVGLPAILLPVFFVASLPLGVYALSSSMWILP  
LVLIISAGTADLQOKRTFMNRVMWMLGDVSPALYVHFLVITYGADLLGFSQTEDA  
PLGLALPMLIPEFLAVSLVLSWLYRFVLPVWRMARPARARRKATEPEQTSERR"  
4651. .5775  
/gene="megDII"  
4651. .5775  
/gene="megDII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"  
/protein\_id="AAC13910.1"  
/db\_xref="GI:10179845"  
/translacion="MTTVMSYLLSEYERERADILDVQKVPASGLILGQSVENFETE  
YARHGIACHGVNGTNAVKLAEVSVCGRDEVVTNSPAAPVLAIDIGARPV  
ADGDEVYIMOTDLVEAAVTPRTAIVPVHLYGOCVMTALRELDNRGLVLEDCAO  
AHGARBRRLAGTMSDPAFSPFPTKULGAGDGAAYVNDDETARLRRLRYGME  
VYVYTRPETHNSRLDEVDAILKRKLTLDAVAGRAVAQRIVDGLADDSGEL  
PVPYDGNHVFYVVYVRRRDEIIKRLDSYDLSLNTSYPMVHTMGFHLVAVSG  
SLPVTERRAGEIFSLPMYPSLPHLDQDRVIEAVREVITGL"  
5822. .6595  
/gene="megDIII"  
5822. .6595  
/gene="megDIII"  
/codon\_start=1  
/transl\_table=11  
/product="daunosaminyl-N,N-dimethyltransferase"  
/protein\_id="AAC13912.1"  
/db\_xref="GI:10179847"  
/translacion="MPNSHTSTSTVAPYERADLYHDFHGRKGTYAEADALVEVA  
RKHTPOAATLLDVACGTGSHVELADSFEVGVDSLAAMLATARNDPGRHLQGM  
RDFSLEDRPDVYTCFSGTGLVLEADRAVANLAGLACGTLLVPEPMPEFTRP  
GMVGADLYTSGDRRTSRMSHTVPAGLPDRTASRMTIHTVGSPEAGIEHFTVEVMTL  
PARAYEAPQPAIGSCSYGHDLFSPLFLGVAAEPRR"  
6592. .7197  
/gene="megDIV"  
6592. .7197  
/gene="megDIV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"  
/protein\_id="AAC13911.1"  
/db\_xref="GI:10179846"  
/translacion="MRVEELIGVFTFPOTFADERGVTGAQOEDFVALGRPLF  
PYAQVSTRSRGVGVGHFTTMSGMAKYVYCAGRPMDFAVDIPRSPFGRALPV  
EYASNSVGLYTPVGMGHLFVSLBEDTTLVYMSGYVPEKRAVHPDPELALPIPA  
DLDLVMSERDRVAFPLREARDGILPDYAAACRAAHRVVRT"  
7220. .8206  
/gene="megDV"  
7220. .8206  
/gene="megDV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/protein\_id="AAC13913.1"  
/db\_xref="GI:10179848"

gene  
CDS  
complement (8228. .9220)  
/gene="megDVII"  
complement (8228. .9220)  
/gene="megDVII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"  
/protein\_id="AAC13914.1"  
/db\_xref="GI:10179849"  
/translacion="MGTTGASRVVGRSALHTSRLMLGTGVNFSGRVTDALRLMD  
HALERGVNCTIDTADYGMRLYKHTRELVGRWPAQGGREETVLATKVSSESRN  
DGELSRHYTAACENSRLRGVNHIDITQTHNIDRAAPMEQWQALBHLVSGKGVY  
GSSNLGSMWHTAAQESAAARRNLGMSHQCILVLAVRPELDVTPAQAQYGVYFAS  
PLHGGLSGVLEKTLAAGTAVKSAQGAQVLLPVRPLVEAYEDYCRRLGADPAVGLA  
WYLSRPGILGAVIGPRTPEQLDSRLAAELITLGEERLELEAIFPAPAVDGPV"  
complement (9226. .10479)  
/gene="megBIV"  
complement (9226. .10479)  
/gene="megBIV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-mycarose glycosyltransferase"  
/protein\_id="AAC13915.1"  
/db\_xref="GI:10179850"  
/translacion="WRVLTLSFARHTHFGQLVPLANALHTAGHDVAVASPELTDVVY  
GAGTLPVLSIDRLFPDISPEAAQVRYTTDIDFARGSELBSWELHGEIETSTF  
VFPVYVNDSSFVDELVEFPAEMDRDLVMEFETPAGAFAACAAHARLLMGSLDLY  
FRSRNLDRCORPADPMDPLGMLTEFAFRGIDYEDDAVQSMVDQLPESRRLT  
GLSYVTRTLPYNGSSVYPMQLTSDGVRVCFGTGSAVIGTSPQEFRTIATLAR  
FDGEIVYTRSGRLPASFPDVRVLDVTPVNNHLLPGCAAVIHGGAGSMATRLHGVQ  
ISVAHENDCVLRQRTNLELAGVFLRPDEVADTLTMOALATVVEDRSHAENAEKLRQ  
ALAAPFAEVVVPVLEALAHQRADR"  
complement (10483. .11424)  
/gene="megBIV"  
complement (10483. .11424)  
/gene="megBIV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/protein\_id="AAC13916.1"  
/db\_xref="GI:10179851"  
/translacion="WTHVTLIGSVGVGSLALREFTTHPLRLAVALVARGSDRPPS  
AGIEHLVLDLEPGRVAQVADVDVVALVAYAGSTWRSAAVPAERVNAIGIMKD  
LVAAILARPQGPAPVTLFASSTQANPAAPSRVYQHKIEARILNQATEDGVADVTILR  
LPATYHSGSGSGGTGRTVAMIRBALAGEPITWMHSGSVARNILHVEDVATFTAL  
HNHRAIVGWTWPSADRARPLGEIFETVVAASVARQGNPAVPVSVPPENAEANDR  
SDPDSIFERTLTGMHPRVPLABGIDRTVAALISTKE"  
12181. .22821  
/gene="megAI"  
12181. .22821  
/gene="megAI"  
/note="polyketide synthase"  
/codon\_start=1

Query Match 29.2%; Score 2606.8; DB 1; Length 47981;  
Best Local Similarity 99.9%; Pred. No. 3.7e-177;  
Matches 2608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6415 CTCGAGCCGATGCTCGGCGCGCGCGGCGCAACGTCGTGACGTCGTGCGGCGGT 6474  
|||  
DB 1 CTCGAGCCGATGCTCGGCGCGCGCGGCGCAACGTCGTGACGTCGTGCGGCGGT 60  
|||  
QY 6475 GGGAGGTCCGCGCGCGGCGGATCAGGAACGATTGCGGATTCGCGAGTCCG 6534  
|||  
DB 61 GGGAGGTCCGCGCGGCGGATCAGGAACGATTGCGGATTCGCGAGTCCG 120  
|||  
QY 6535 ATGACCGTTGACCCGATCCCATATGCGCTCTCCCGTATGTCGTGGCGGTCG 6594

121 ATGACCTTGAACCGGATCCCCATACGCTCTCCCGATGATGCTGGGGCGGTCCGTCCGG 180  
QY TACCGCCCGGACTGACATTTGTGATGATGACACCCCGCCCACTGATGAGGCTCCGCCCGCA 6595  
Db TACCGCCCGGACTGACATTTGTGATGATGACACCCCGCCCACTGATGAGGCTCCGCCCGCA 181  
QY CCGGAGAGAGGTCCCTGCAACAATTCGCGGTGACCGGTCCCGCGGTCCGTGAAACGAGG 6555  
Db CCGGAGAGAGGTCCCTGCAACAATTCGCGGTGACCGGTCCCGCGGTCCGTGAAACGAGG 241  
QY GTGAGAGACCCCATCATTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6715  
Db GTGAGAGACCCCATCATTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301  
QY TGAACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6775  
Db TGAACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361  
QY GACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6835  
Db GACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421  
QY GAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6895  
Db GAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481  
QY CATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6955  
Db CATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541  
QY CGTGCATCATCAGCTGCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 7015  
Db CGTGCATCATCAGCTGCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 601  
QY GAAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7075  
Db GAAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661  
QY GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7135  
Db GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721  
QY GAAATCCGATCAGCTGCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 7195  
Db GAAATCCGATCAGCTGCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 781  
QY GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7255  
Db GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841  
QY TCAAGCTTCAATTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7315  
Db TCAAGCTTCAATTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901  
QY GAGTTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7375  
Db GAGTTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961  
QY ATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7435  
Db ATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1021  
QY GGAAGGAGAGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7495  
Db GGAAGGAGAGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1081  
QY GCGAGAGAGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7555  
Db GCGAGAGAGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141  
QY CAGGTGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7615  
Db CAGGTGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7674

1201 CAGGTGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
QY ACCGAGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7675  
Db ACCGAGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1261  
QY CTGAGTAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7735  
Db CTGAGTAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1321  
QY TGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7795  
Db TGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1381  
QY TGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7855  
Db TGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1441  
QY ACCGAGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7915  
Db ACCGAGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1501  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7975  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1561  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8035  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1621  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8095  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1681  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8155  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1741  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8215  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1801  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8275  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1861  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8335  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1921  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8395  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1981  
QY GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8455  
Db GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2041  
QY CAAAGAGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8515  
Db CAAAGAGAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2101  
QY GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8575  
Db GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2161  
QY GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8635  
Db GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2221  
QY GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8695  
Db GGTAGTAGGCGATGCAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2281





TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9 =  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37812.1"  
/db\_xref="GI:13938933"  
/translation="MNVLTSGAFLGSAVTHALDLPVRLVVARREVVSAGVADYE  
THRVLDLEPGLAEVADAVPFAQIIGTSGMRISBDVVAERTNVGLDILAV  
LSRSPHAPVVPFGSNTQVGTATGRTIDSDHDPGIVDRQKHGEQLKERTAG  
AIRATSLRLPVPVPGPAAGADRGVMTIRALTCQPTLMHGDGIRELLIYVDA  
ARAFTVTHADNLAGHRLIGTRSPCLSEVQVARSARTGSDPVEVSPVPPA  
HMDSDLRSEVVPARTAVTGMATVTVMAEADRVTAALAPRRAPSPSS"  
complement(8228..9220)  
/note="unnamed protein product; megB11-1(megDV11),  
TDP-4-keto-6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37813.1"  
/db\_xref="GI:13938934"  
/translation="MGTGAGSARVVRGSAHTSRMLTGTVNPSGVYDDALRMD  
HALRGVNCIDPADYIGMRLYKHTSELVGRMFAQGGREEVTLATKVSSESRVN  
DGLSARHIVACENSLEKRLGVHDHIDYQTHIDRAPMDEVMQAAHVLGSGKGVY  
GSSNLGSMHIAAQAESAARMLGMSHOCYMLAVRPELDVLPAOAGVGFAMS  
PLHGILSGVLEKAGTAVKSAQRAQVLLPAVRPLVEYEDYCRRLGADPAVEGLA  
WVLRPGILGAVIGPRTPEOLDSDALRAELTIGEBELRELEAFPAVAVDPVP"  
complement(9226..10479)  
/note="unnamed protein product; megB1, mycosyl  
transferase, mycrose glycosyltransferase; SEQ ID NO: 11=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37814.1"  
/db\_xref="GI:13938935"  
/translation="MEVLLTSFARHTFQGLVPLAMALHTAGDVAAGPBLTDDVY  
GALVTSPGLSDRLPDISPBAQAQVHYRTTDLFARGBELSMWELGIBERTGF  
VFPVNNDSFVDELVEFAMDWRBDLVMEPTFAGAVAAACAARALMGSDLTGF  
FRSRQDLRGORPADRPDLGMLTEVARGFLDYSDLAQVOMSDVDDPESFRLT  
GLSEVHTRLPYNGSSVVPQMLTSDGVRCCTGCVSALGITSNQBELTATLAR  
PGEIYVTRSGDLPASVPDVRVLVDVPMNITLPGCAAVIHGGASQSMALHGVPO  
ISVAHEMDCVLRQETAEAGVFLRPEDVDATLQWALATVEDRSHAMENKLOE  
ALAAPPAEVPVRLBALAHQHRAD"  
complement(10483..11424)  
/note="unnamed protein product; megB1V, TDP-hexose  
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 12 = translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37815.1"  
/db\_xref="GI:13938936"  
/translation="MTRHTYLLGSGFVSALLRETTPLRLKRAVARTGSRDQPGS  
AGIEHLRVLDLEPGLAEVADAVPFAQIIGTSGMRISBDVVAERTNVGLDILAV  
LSRSPHAPVVPFGSNTQVGTATGRTIDSDHDPGIVDRQKHGEQLKERTAG  
LPAIYGHSPSGQGTGRTVATMIRALAEPLTMHGEVIRRLALBEDVATFTAL  
HNHEALVGVMTSPSADBARPLGEIIFETVAASVAROTGNAPVAVSPPEMNEANDFR  
SDPDSTERRTLTGMRVPLAEGIDRTVAALISTGE"  
12181..22821  
/note="unnamed protein product; megA1, SEQ ID NO: 13=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37816.1"  
/db\_xref="GI:13938937"  
/translation="MVDVLDLGTTRTHPGLPFPMPPLCGHNEBELBARAQHAYLE  
GISDSDVVAAGALAEETRAODGPHRAVVAASVTELTALALAAQGRPHSPVVRVA  
RPTAPVFTLPGGAQMPGMAFRLAESVFAAAMACRAFEVLDMSLTELVDPE  
HLRREVVOPALPAVQTSLLAALMSRSGVPDAVIGSISGLAAABVCGADVAARA  
AALMSREVPVLRGDMAAVLSAPALRAVERMDDDVPAGNGRSVTLGCAAPRI  
ABRVAEIAQGVRAQVYNTMSMAHSAQVDAVAGKMSALTWRPGSDVPTYAGLTG  
RIDPTELDHMRKFRFLFVRDEATRAVLELQPTFISSPHVLAASLQOTLDEVG  
SPAPVFTLQRDQGGRLRLFLAVQAATGVVDVMTGAAPGVPTGLPAPVAVETDEG  
PSTGEPMAPDHLARLRLLEIAGAEALAGREVDAATFRREIGLDSVAVOLATLA  
TATGRDLTAMLYDHTPHALTALRQGEERGRGETAHPTEAPRDPVAVVMAAC

RLPGVTSPEEFWELLABGRDAVAGLPTDRGMDLSLPHDPTRSGTAHORAGFLTG  
ATSPDAPFGLSPREALAVPPOORTLLESMWELREAGIPSTLSRTSGVFLIQ  
KXGPRABRGEVGEVLTMTGTTTSVSGVAATYLGEGRAISDTTRACSSSLVAVHLAC  
OSLRREESTMALAGVTVPTPGLMVDPSKMSLSLADGKASKASALADGEMGACH  
LLBERLSDRRHGPVLAVIRTAVNSDASNLSAPNGAQRVIRQALAESGLTPEH

Query Match 29.2%; Score 2606.8; DB 6; Length 47981;  
Best Local Similarity 99.9%; Pred. No. 3.7e-177;  
Matches 2608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

6415 CTGACCCGATGCTCGCGCGCGGCGGCGCAACAGTCGTGACGTCGTGCGGCGGT 6474  
1 CTGACCCGATGCTCGCGCGCGGCGGCGGCGCAACAGTCGTGACGTCGTGCGGCGGT 60  
6475 GGAAGGTCCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6534  
61 GGAAGGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
6535 ATGACCGTGAACCCGATGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6594  
121 ATGACCGTGAACCCGATGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
6595 TACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6654  
181 TACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
6655 CGGGAAGAGTCCGTCGGAACATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6714  
241 CGGGAAGAGTCCGTCGGAACATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
6715 GTGCGAGCAACCGGATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6774  
301 GTGCGAGCAACCGGATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
6775 TGAACCGGTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6834  
361 TGAACCGGTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
6835 GACCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6894  
421 GACCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
6895 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6954  
481 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
541 CATGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
6955 CATGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7014  
541 CATGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
7015 CGTCGACATCACCGTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7074  
601 CGTCGACATCACCGTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
7075 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7134  
661 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
7135 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7194  
721 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
7195 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7254  
781 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
7255 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7314  
841 GGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
7315 TCACAGCTTTCATCCGAGCGGCTTCTATGAGAGTTCGAGTCGAGCGGCGGCGGCGGCGG 7374  
901 TCACAGCTTTCATCCGAGCGGCTTCTATGAGAGTTCGAGTCGAGCGGCGGCGGCGGCGG 960



OY	7275	GAGTTGCAAGATGGCCCGGGGGTCTTCACTGGGGGGTTGGGTGCACGGGCAGATCTGACTCG	7434
Db	961	GAGTTTGCAAGATGGCCCGGGGGTCTTCACTGGGGGGTTGGGTGCACGGGCAGATCTGACTCG	1020
OY	7435	ATGCTCTCTGTCCGAGCCGAGCAAGACCCCTTGACCTTGATACGAACGGTTTCGGGGCCGC	7494
Db	1021	ATGCTCTCTGTCCGAGCCGAGCAAGACCCCTTGACCTTGATACGAACGGTTTCGGGGCCGC	1080
OY	7495	GACACGGGGACCGGTACCGCCAGTTCGGGGCCGGAAACGTGGGTGGTTCGGGTACACACCGGACCGCC	7554
Db	1081	GACACGGGGACCGGTACCGCCAGTTCGGGGCCGGAAACGTGGGTGGTTCGGGTACACACCGGACCGCC	1140
OY	7555	GCCGAGGTGCTCGCGCAATTCGGGGGCTTTCACCAACGGGCCGAGCCGACGCTGCACGGGTGGATG	7614
Db	1141	GCCGAGGTGCTCGCGCAATTCGGGGGCTTTCACCAACGGGCCGAGCCGACGCTGCACGGGTGGATG	1200
OY	7615	CAGGTGACCCACTGCGCCCGAGCGACTCTCTGGAGCCGAGCCCTTCCGGAGAGTTCTACGCCGCG	7674
Db	1201	CAGGTGACCCACTGCGCCCGAGCGACTCTCTGGAGCCGAGCCCTTCCGGAGAGTTCTACGCCGCG	1260
OY	7675	ACCGAGAGACGCGCGGTCCGTGACAGTGGACCGCCACTGGGCTTCAAGACGGGTTCGCCAGG	7734
Db	1261	ACCGAGAGACGCGCGGTCCGTGACAGTGGACCGCCACTGGGCTTCAAGACGGGTTCGCCAGG	1320
OY	7735	CTGGGTGACCGAGCTGGGGGTTCGCGCTTTCGATCTGCGAAACGACTTTCGCCCGGAGAGTCCCG	7794
Db	1321	CTGGGTGACCGAGCTGGGGGTTCGCGCTTTCGATCTGCGAAACGACTTTCGCCCGGAGAGTCCCG	1380
OY	7795	GTCGTGACGCTTCGGTACCGCGCCCGCACTCAAGGGCGTGGACCCCGACCGTCTCCGGTCC	7854
Db	1381	GTCGTGACGCTTCGGTACCGCGCCCGCACTCAAGGGCGTGGACCCCGACCGTCTCCGGTCC	1440
OY	7855	TGGACCTTCGGGCGACCCCGGGTATGCTTGGACCGCCCAAGGTCAACCCCGGCAACAGCTGGCGGTG	7914
Db	1441	TGGACCTTCGGGCGACCCCGGGTATGCTTGGACCGCCCAAGGTCAACCCCGGCAACAGCTGGCGGTG	1500
OY	7915	ACCGAACAGGGCGCTACACCGCCCTCGACAGAGATCGACCGCGGTCAACCGCGCGGTCCGGAGCGCC	7974
Db	1501	ACCGAACAGGGCGCTACACCGCCCTCGACAGAGATCGACCGCGGTCAACCGCGCGGTCCGGAGCGCC	1560
OY	7975	GCGGTGCTGTGGGGGGTGGTGGCGGAGCTGGCGGCAACAGCGTGGGCAACGCGCTCTG	8034
Db	1561	GCGGTGCTGTGGGGGGTGGTGGCGGAGCTGGCGGCAACAGCGTGGGCAACGCGCTCTG	1620
OY	8035	GCGGTCAACCGAGCTTCCCGAACTTGGCGGCAAGACTTGGCCAGACACCCGGAACCCGCAAC	8094
Db	1621	GCGGTCAACCGAGCTTCCCGAACTTGGCGGCAAGACTTGGCCAGACACCCGGAACCCGCAAC	1680
OY	8095	GCTGTGTGTGACCGGAGAGTGTCCGCGACAGAGTCCCGCGCGTCCACTTGGAAACGCGGCAACCGCC	8154
Db	1681	GCTGTGTGTGACCGGAGAGTGTCCGCGACAGAGTCCCGCGCGTCCACTTGGAAACGCGGCAACCGCC	1740
OY	8155	GCGTCCGACCCGCGCGGTGGCGCGGGTTCGACGTCCTCCGACCGGTTGGCGAGGTGACAGTGGTC	8214
Db	1741	GCGTCCGACCCGCGCGGTGGCGCGGGTTCGACGTCCTCCGACCGGTTGGCGAGGTGACAGTGGTC	1800
OY	8215	GTCGCGCGCGGCAACCGTGAATCCCGAGGTCCTTCACTGATCCCGACCGGTTTGAAGTGGAC	8274
Db	1801	GTCGCGCGCGGCAACCGTGAATCCCGAGGTCCTTCACTGATCCCGACCGGTTTGAAGTGGAC	1860
OY	8275	GCTGGCGGCGAAGCGCCGAGATCCTGTGTCGCCCGGCGCGGCTGGCCCGGACCGGACCTTCGAC	8334
Db	1861	GCTGGCGGCGAAGCGCCGAGATCCTGTGTCGCCCGGCGCGGCTGGCCCGGACCGGACCTTCGAC	1920
OY	8335	GCCCTGTGTGACCACTTCGCGCAACCGCGCGGCTTGGCGGGCCGCGCGCCCGGTGTTGCCCGG	8394
Db	1921	GCCCTGTGTGACCACTTCGCGCAACCGCGCGGCTTGGCGGGCCGCGCGCCCGGTGTTGCCCGG	1980
OY	8395	CTGTCTCCGTTCCGGGCGCGGTGATTCAGAGGAGCTCGGTCAACCCGTGCCCCGTGTGTTCAGG	8454
Db	1981	CTGTCTCCGTTCCGGGCGCGGTGATTCAGAGGAGCTCGGTCAACCCGTGCCCCGTGTGTTCAGG	2040

Oy		8455	CGTGGCCCGGTGAGGAAAGAAGACGGGTCGTGTTTCACTGAATGGCTGT	8514
Db		2041	GCTTGCCCGGTGAGCTGTMAAGGAACAATGCCGTCGTGTTTCACTGAATGGCTGT	2100
Oy		8515	CAACAGGCATCTGTTCGGGCTGTCCGCTTCGCAAGCGCTTCCAGCGGCGCAGACA	8574
Db		2101	CAACAGGCATCTGTTCGGGCTGTTCGCTTCGCAAGCGCTTCCAGCGGCGCAGACA	2160
Oy		8575	GGTACGGGTCGTGGCTTCGCCGAGCCCTGACCAGACAGTCAACCGATCCGATCTGACCCG	8634
Db		2161	GGTACGGGTCGTGGCTTCGCCGAGCCCTGACCAGACAGTCAACCGATCCGATCTGACCCG	2220
Oy		8635	CGTGC CGGTGAGTACGACGCTGGAACCTTGTGAGATTGGCAAGCCCAAGCGGAGCAGACAT	8694
Db		2221	CGTGC CGGTGAGTACGACGCTGGAACCTTGTGAGATTGGCAAGCCCAAGCGGAGCAGACAT	2280
Oy		8695	CGTGCAGTACATMGCGGACCTTCGACTGGGTTCAGCACAGAGCACACACATGTTCTTGGGA	8754
Db		2281	CGTGCAGTACATMGCGGACCTTCGACTGGGTTCAGCACAGAGCACACACATGTTCTTGGGA	2340
Oy		8755	CGACCTCTTGAGGATGACAGACCAACCTTCAACCAGACTTCTTGCGCCTGATGAGCCCCGA	8814
Db		2341	CGACCTCTTGAGGATGACAGACCAACCTTCAACCAGACTTCTTGCGCCTGATGAGCCCCGA	2400
Oy		8815	CTCGCTCATTCACGCGAATGATTCAGATTCTGCGGCTCCGAGGTCGCGAACGTATGATGCTG	8874
Db		2401	CTCGCTCATTCACGCGAATGATTCAGATTCTGCGGCTCCGAGGTCGCGAACGTATGATGCTG	2460
Oy		8875	GGAAGCCGTGACTTTCGCGCCCGGATTCGCGAGCCCGGATTCACCGAAACCCCGACGCGCG	8934
Db		2461	GGAAGCCGTGACTTTCGCGCCCGGATTCGCGAGCCCGGATTCACCGAAACCCCGACGCGCG	2520
Oy		8935	GATGCTGTGGGATCCGACGTCGCAACCCGGGACCCGACAGACTTCTTGAGATGCTGTGGC	8994
Db		2521	GATGCTGTGGGATCCGACGTCGCAACCCGGGACCCGACAGACTTCTTGAGATGCTGTGGC	2580
Oy		8995	CCACCAAGAGGTGAGACACCGGAGAGATCC 9024	
Db		2581	CCACCAAGAGGTGAGACACCGGAGAGATCC 2610	
RESULT 8				
AY623658				
LOCUS	AY623658	61845 bp	DNA linear	BCT 16-FEB-2005
DEFINITION	Aeromicrobium erythreum putative transcriptional repressor and putative dehydrogenase/reductase genes, complete cds; erythromycin biosynthesis gene cluster, complete sequence; putative oxidoreductase and lipN genes, complete cds; monamine oxidase gene, partial cds; and unknown gene.			
ACCESSION	AY623658			
VERSION	AY623658.2	GI:59723038		
KEYWORDS	.			
SOURCE	Aeromicrobium erythreum			
ORGANISM	Bacteria; Actinobacteriales; Actinomycetaceae; Bacteroidetes; Nocardioideales; Aeromicrobium.			
REFERENCE	1 (bases 1 to 61845)			
AUTHORS	Brikun,I.A., Reeves,A.R., Cernota,W.H., Jau,M.B. and Weber,J.M.			
TITLE	The erythromycin biosynthetic gene cluster of Aeromicrobium erythreum			
JOURNAL	J. Ind. Microbiol. Biotechnol. 31 (7), 335-344 (2004)			
PUBMED	15257441			
REFERENCE	2 (bases 1 to 61845)			
AUTHORS	Brikun,I.A., Reeves,A.R. and Weber,J.M.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-MAY-2004) FermaLogic, Inc., W. Campbell Park Drive, Chicago, IL 60612, USA			
REFERENCE	3 (bases 1 to 61845)			
AUTHORS	Brikun,I.A., Reeves,A.R. and Weber,J.M.			
TITLE	Direct Submission			
JOURNAL	Submitted (14-FEB-2005) FermaLogic, Inc., W. Campbell Park Drive, Chicago, IL 60612, USA			
REMARK	Sequence update by submitter			







OY	6226	CACATCATGACCCGAGAAACCCGATCTCGGTTTGACGATGATGGATTGCATCCAGTCCCG	6285
Db	9315	CACAGACAGGGCCGAGGAACCCGACTCCGGCTCGACGAGATTCGGCTGCACACTCGGC	9374
OY	6286	TCGCCATTCGAAGTTGCTCGCGACGTGCAGGACCTTCGATGAGAGAAGACCCCGAGTC	6345
Db	9375	CCGCCACCCGGTGTTCGTGCAGACGGCCCAACCCCTTCGACCGAAGAAACCGGCCGAAATC	9434
OY	6346	GTGCGCCACCGCAGACCGTCTCTCCGGGTGGAACGACGAGTTCCATGTGCTGAAGGCAC	6405
Db	9435	GTGCGCGATTCGCTCCGTGCGTGCAGGCTCCAGTCTCAAGAGTCTGGCGAAACGAGT	9494
OY	6406	TCGGTGCACCTTCGAGCCGATGCTCGGCGCGCCGCGGAGCCACACGATGCTGAGCGTC	6465
Db	9495	CCGCTTCGACCTTCGAGGTGCAGACCGCGGACCGCTTCGGCGAGCACCGACGACGAGCTC	9554
OY	6466	G 6466	
Db	9555	G 9555	
RESULT 9			
AB088224/c		210614 bp DNA linear BCT 11-JUN-2003	
LOCUS	AB088224	Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.	
DEFINITION	AB088224		
ACCESSION	AB088224.1	GI:30698345	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			

implemented at <http://www.nih.gov/jp/~jnm/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, gtc, ctg or tgg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lanthacin, lankamycin and an unknown type II polypeptide, a carotenoid biosynthetic gene cluster, many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

Location/Qualifiers

1. .210614

/organism="Streptomyces rochei"

/mol\_type="genomic DNA"

/strain="7434AN4"

/db\_xref="taxon:1928"

/plasmid="pSLA2-L"

/note="Linear plasmid"

1. .1992

/note="Left terminal inverted repeat, TIR-L, shows 99.4 % (1991/1992) sequence identity to TIR-R (complement 1208623.210614) "

683. .2188

/note="N-terminal sequence is almost identical (435/437) with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L

ORF1 (501 aa)

similar to A0590463 Streptomyces coelicolor putative helicasae, SCPI.136 (879 aa); homology is seen until the inner end of TIR-L

/codon\_start=1

/transl\_table=1

/product="putative helicasae"

/protein\_id="AC76459.1"

/db\_xref="gi:30698346"

/translation="MSTTSRTDQREAAQAEVADVAVRALPRLPLAPEGGLTQVIMATGSKRTVRAAEKLRAGKRVLVLPSSDLLQTETAMEAGTGMIVSSLRGDAVAFPTDVEELIDMVRPFDKVTVFATYASIGLTERARGGIPGMDLIVDEARHTSGRLGKPAVWMDTRTIPSLRLYMTATPRLMOMDEBAEGAPGELVMSMDPDGELGALRKSPITSEADIDRGICAPQVYCVDTTPQLOAAOLGVGRSDVGRGAAALOTAPLKKASSENFRTLVFHMVMEKAEAPFAGLDPVAKRLHAAGPGCIYPTTIANMLCGHKKFRRRLRILGFTSGIATIDGTVEKFGISVKGEGVDTRCEDSVYMAVRKSMPLVOAVGRARILIPGGQKASLVLPVLLDPGETADNMLTSRPNYGALFRDAPLLTGHGEEGFAAFPEGEFVRPVLYGVGAQRGASAGPASRSITPSAMSSISIMRAP"

complement (2315. .3595)

/note="ORF2 (426 aa)

similar to AEO04736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)"

/codon\_start=1

/transl\_table=1

/product="putative secreted protein"

/protein\_id="BACT6460.1"

/db\_xref="gi:30698347"

/translation="MRRKPRALIGIPGLVLAVLAPAAHASNADIVYGTMAAPPTAAPASDEVDDTLKQTVHLISVAGSDILVRFFNEIGTSPLTIGVHARPAAGCATVADGDRVAVRFGRRGADLTPVGTSTSMRPIISGVSDRAGATGSDSVLTGDSIGDEHTYQHNVAAGDVTGAPADLTTPVGTSTSMRPIISGVSDRAGATGSDSVLTGDSIGDEHTTIDARRRPDLAERLRDGLACTGVNNAIGTGNRLRPDPBPGSAASPAAYFEB SALKRGLVDVIGQPARAVTVLLGVNDLQGGIAAPASDEVTAELIAGRLQLERAH EHGKLYCATITPPAGDITGYFTPRRRAVRQOVNDWRTSGAPDTVIDFPAVAVLDPARPDLHLPAYDGGDGLHPNDAGMAAARAFPLDSLR"

complement (3659. .4645)

/note="ORF3 (328 aa)

similar to Y00459-2 Streptomyces griseus regulatory protein, StrR (350 aa)"

/codon\_start=1

/transl\_table=1

/product="StrR-like regulatory protein"

/protein\_id="BACT6461.1"

/db\_xref="gi:30698348"



CDS  
/translation="MTAALAGNDIETRSIPISLVGSPRLCGEDAEHVAIADTD  
EALPILIVERSTRIIVDMKRVLAAXKAGHTIEVLFDGAEEAFILAAKSMHTGL  
PLSROBRBAORIIAQMHPHLSRAVAGTAGIKATYAAALRAAGTAPRAORGD  
GRIRPLDTGGRKAALLAQRPAQSTREYARAGISPAIASVRRRLAGSPVER  
NTGAPAGGTGSRATPGGAPEPVSVPIRPEDSPILVRDPSLRHKSGRRL  
LRLQCGAVERTLMAAQTPPHCTDLVAELREYADLVAEPAREVARDG"  
complement (4922) .5812)  
/note="ORF4 (296 aa), lankacidin biosynthesis protein  
similar to AE004625-4 Pseudomonas aeruginosa  
pyrroloquinoline quinone biosynthesis protein B (304 aa)"  
/codon\_start=1  
/transl\_table=11  
/product="pyrroloquinoline quinone biosynthesis protein B"  
/protein\_id="BAC76462.1"  
/db\_xref="GI:30698349"  
/translation="MILITRAGAGFPFNNCAALCARRGELPARSGCAVSGDG  
RDMWLNASPDITQLLAAPALTPGPRIIPVRGVLITAEVDHLAGLVAGTGL  
TVYAAPVRGALSAELPVRLDRVAPWMDRATAPGPAVAGLITVAPVTXAK  
YAHAPDPAWVCAVRIEDPATGALVYACLATPDGPDILASATCALDPSFA  
GELGATSSAGAGOSLMGHLFVAGPGSLAALRHGRLRIYTHANNPLDPSGA  
HAAREAGVEVLPDGESELYT"  
complement (5815) .6909)  
/note="ORF5 (364 aa), lankacidin biosynthesis protein  
similar to AJ277117-5 Gluconobacter oxydans putative pqsB  
protein (359 aa)"  
/codon\_start=1  
/transl\_table=11  
/product="pyrroloquinoline quinone biosynthesis protein E"  
/protein\_id="BAC76463.1"  
/db\_xref="GI:30698350"  
/translation="MADPAVGAAGNLIIEIHRCPHCPYCSNPLIEVREAEITCEO  
MTDILTOARELGVGMHPSGSEPLARDPLDVGHARBAVYVITSGVGLTAEEAH  
DIARCGDVHOLSLDAPDPAAGQALAGAVHTKAAARVTAAGIPUTNYITHGN  
IDRTGWMVDLADLGDRIELANTQYTGMLRNALRAALAEAREVNRAR  
AGSELVYAADYDPRKPCMDGWSITDITVPADVLPCEPAVATITLPSNARL  
PLSEIVASRSPFAYRGTMREPCREHADHGSCRCQAFQLTGDAATDPACGL  
SPHRSIVDAALAEITDGPVAFVPRGFPVA"  
complement (6902) .7174)  
/note="ORF6 (90 aa), lankacidin biosynthesis protein  
similar to AL603642-200 Sinorhizobium meliloti putative  
pyrroloquinoline quinone biosynthesis protein D (98 aa)"  
/codon\_start=1  
/transl\_table=11  
/product="pyrroloquinoline quinone biosynthesis protein D"  
/protein\_id="BAC76464.1"  
/db\_xref="GI:30698351"  
/translation="MTGLPEPTVPRLPQVRLTRDPARGELALLPVRVWLANDTAAY  
LAHCDGTSLAGIVERLAEYEGVSAEDVRELLRLAQRVVDLHG"  
complement (7171) .7899)  
/note="ORF7 (242 aa), lankacidin biosynthesis protein  
similar to AE004625-5 Pseudomonas aeruginosa  
pyrroloquinoline quinone biosynthesis protein C (250 aa)"  
/codon\_start=1  
/transl\_table=11  
/product="pyrroloquinoline quinone biosynthesis protein C"  
/protein\_id="BAC76465.1"  
/db\_xref="GI:30698352"  
/translation="MSMSVTRVPAWSEAEFRQLALHLESSYWDHPPHRRMEGL  
DSGELALMANRWYQRCQKDAIVANCPLEVRQWLSRVYHDGADACGAGEX  
WRLAEAVGRLRDEVDRLVLAGTAPADAVDFPARRPAAAGTLEPSGL  
AHRGLREHPVIAEGEFYFARIEVVGPEGRSLDLVARRAVSRQEOACVRLA  
PFCRVINAVLDSIDVHTGNGATRS"  
complement (7926) .8030)  
/note="ORF8 (34 aa), lankacidin biosynthesis protein  
similar to AL603642-197 Sinorhizobium meliloti putative  
pyrroloquinoline quinone biosynthesis protein A (31 aa)"  
/codon\_start=1  
/transl\_table=11  
/product="pyrroloquinoline quinone biosynthesis protein A"  
/protein\_id="BAC76466.1"  
/db\_xref="GI:30698353"  
/translation="MRTSGKELPKAKAMHRPDTIDTGMETVATYPSR"  
8420. .10057

/note="ORF9 (545 aa), possible lankacidin resistance  
protein  
similar to AL133424-19 Streptomyces coelicolor putative  
ABC transporter ATP-binding protein (544 aa)"  
/codon\_start=1  
Query Match 16.9%; Score 1513.6; DB 1; Length 210614;  
Best Local Similarity 59.5%; Pred. No. 6.9e-100;  
Matches 2563; Conservative 0; Mismatches 1736; Indels 13; Gaps 5;  
1453 GCCAGCTGTGGGGCTGCCCGGATGACCAAGAGATTGCTGTCGGCGCC 1512  
92581 GACTGTGTGTGGAGCGGTACAGTTTCGGGGCCATTCGGGGGGGTGGGCG 92522  
1513 CTGTGACATTCAGATGACAGACCCGACCGATCCGCGCTGTGACAGCATATGACAG 1512  
92521 GCGGCCACGCGCGTGTGTCTCCGCGAGAGCACTTAAC--GCTTACTTCCCGGAGCG 92465  
1573 GTGCTGAACCGGCTCACTTCCTTACCTGTCGACAGGTTGGGAGCGGGCGGACCC 1632  
92464 TTCTTGACCGGCTTCGCTCCAGACAGCCCGCTGACGGGCTCCGCTGCGGAGTGG 92405  
1633 CGGAGACCTGATCTCCGCGCTGTGTGCTGCGCGAGTCAAGCGGCGACCTTCAGAC 1692  
92404 CTGGCGGACGCGCGCGGGGGCCACGGGGGTGCGTTGACAGAGAGCTGTGTGCGCAG 92345  
1693 GTGAGGCGGCACTTCTTCACAGCGTTGCTGTGCGGGGCAATCAACCAACCGTC 1752  
92344 TGGAGGCTGACCTGTGCTGCCGAGCGGCTTCGCGCTGACCTGTGCTGCGACGTC 92285  
1753 CTGCTGGGCAATCTGCTCCGACCTTCGACAGACACCGGAGTCACTGACGCGCGCGCC 1812  
92284 CTGCGGTAGTCCGCTTCAACAGCGCGCGCGCTGTGCTGCTGCGCAAGCGCGGG 92225  
1813 GAGAACCGGCTGTGATCCGCGCGATCATGAGAGATGTGTGCTTCCGCCCGCTTC 1872  
92224 CCGCGCGCGGGTGTGCGTGAAGAGCGGGCTCAACGGGAGACCGCTTCAACCGCGCC 92165  
1873 CCCAGATGAGCGGACCAAGACAGAGGCGACACCGTGTGCGGGTCAAGATCCGGCC 1932  
92164 GCGGTCTTCAACGTTACTCTCGGGTGTGCGGAGGTGACGCGGAGATCTGTCAC 92105  
1933 GACGTATGATCAACCTGCGGTCTGTGCGCAACCGGATCTCCCTGGCGCATCCCGAC 1992  
92104 GCGAGCGCGCGCGCTGTGGGGGTGTGCGCGCTTCGCGACACCTCCGGTGTGCGC 92045  
1993 CCGGACAGTTCGACCGCTGCCGCAAGATCGGTGTGCGCGCGAGCTCTTCGGGAC 2052  
92044 TTGCGCCCGCTGAGCGTCTGTTCGAGAGGTGCGCGCGGTGTCAACAGGTGCGCTTC 91985  
2053 GCGGTGACCTTCTGTGTGCGTCCCGCTGCGCGCGCTGAGAGACAGGTGCGCTGAG 2112  
91984 GCGACCTGTGACAGCGCTGTGACACAGGGGTGCGGAGGTCACTGTGCGCTGCTGG 91925  
2113 GAGATATGCGCCCGGTAGGTGATGATGCGCGCTGACCGGAGAGACAGACGCTGCTAC 2172  
91924 GACACCGTCTCGGGGCGGAGAGACAGCGCGCTGCGCGCGCGCTGTCTTCGCGGAG 91865  
2173 TTGCAACAGATGCTCTCGGACCGCGACCTTCGCGGTCTGTGCGCGGCTCAACCGCGCC 2232  
91864 GCGAGGTGACCGCGGAGGGCTGTGCGGAGCAAGCTGTGCGCGGTCTGCCCAACCTTCC 91805  
2233 GAGTCCGCTGAAC--CCTTGGCTTCGACCGCGCGGAGNNNNNNNNNNNNNNNNNN 2290  
91804 TTCTTCGACGGGGCGCGCGCTGCGGGGCGAGGCGCGCGGAGACAGCCCGACGAC 91745  
2291 NNN 2350  
91744 GTGTGTGCGGTCTGAGAGAACTACCGGACGAGCAAGGGCGCGGTGCGGTGCGGGCG 91685  
2351 CACNCCMCGCGCGCGCGGGGCGGGTTCAGCGCTTTCAGACGGCGCGGATTCAGCGCCCGCA 2410  
91684 GCGGCGGGCTGAGCGCGGGGCGACGCTACGCTGAGAGGCTCGCGCGCCAGCAGCGCCCGG 91625

QY 2411 TGCTGATCCACCCCGTGGGAGAGCTCCCGTTGAGTTGATTGAGCGGGGGGTGACCGGAC 2470  
DB 91624 TACCCCCCAAGCGGTGGGAGAGCTCCGTTCTCATCTGATTCAACCGGGCCGTACACACC 91565  
QY 2471 TGGTGGAAACCGGTGAGGAGAAAGACTCGTCCCGCGCTGGCGAGTGAATGCTGGCGCGAC 2530  
DB 91564 TGGTGGAAATCGGTGAGGAGAAAGACTCGTCCCGCGCGCTGGCGAGTGAATGCTGGCTC 91505  
QY 2531 GTTACGAAGTGTGGAGAGACCAAGTGCAGAGCTCCGGTTCGGGGGTGACCCGGCCGACG 2590  
DB 91504 TCGACGAACCGCGGAGACGACTCGTGGAGGCTCTGTTCCGGGGGTGACCGCGTCCGAGATG 91445  
QY 2591 TAGGGGGTGGCCCCCGGACGCGGGGAGAAACCGGCTCCCGGTTACAGGTAGACGTTCCG 2650  
DB 91444 TAGGGGTCCGCGCGCGGACGAGCGGGAGAGCCGACTGCGGTTAGAGTTACACTTCGCC 91385  
QY 2651 AGCAGTTCGATCTGCACCCGACACTGCGGATGGGCGGTGGGCGCATCTGTTCCGGCGG 2710  
DB 91384 AAAAGTTCGACTGCAACCGGATCTGGGGGTGGGAGAACCGCGCATGGCCGCGGGGTG 91325  
QY 2711 ATCCGCAACAGCTGGGCGTGCACACCCGAGCGAGGCTCTCCAGCGCTAACCCAGGTG 2770  
DB 91324 ATGGCGGAGCGCGGCGTTCACCCCGCTGGCAGGCTCTCAGGGCGTACCCAGATCG 91265  
QY 2771 GTCTGCAATGCCCGGGGTCCGCTCGGCGGGGAGTGTGACGAACCGGGGCGAGGCGTCCCG 2830  
DB 91264 ATGTGACGAGGGGGGTCCGCTCCCGCGGTACGAGTGAAGTCCCGAGCGCGTCCGCC 91205  
QY 2831 AACTCGGTCCGTTCCGCTTCGAGCAACCTGCGGTGTCGACCGCTGTAGTCTTCGCGG 2890  
DB 91204 AGCGGGGGCGCTCGTGGCGCGGAGCGCGCGGTGTCGCGCGCGTGTAGCTTCGCGG 91145  
QY 2891 AGCGTGAAGAAAGTGCAGCGCGCGGTGGCGGCTGGACTGTTGAGTTCCGCGATGAAGTCG 2950  
DB 91144 ACCGTCAAGAAAGTGCAGAGAGAGCGGCGGGGAGCGCTGTCCAGGCGCGGACAGAAAGTCC 91085  
QY 2951 ACCAGTGAAGAAAGCGGTGGCGCGCGCGCGCGAGAGTGTGTTGAGGCGGAGCGG 3010  
DB 91084 AGAAGTTCATGAGCGGTGCGCGCGCGCGCGAGAGAGTGTGTTGAGGCGGAGCGG 91025  
QY 3011 ACCGTGCGCGCGCTTCGCGCGGACATCCGACAGAAAGTCCCGAGGTTCTTCGACCGCT 3070  
DB 91024 GTCCGGGGCGCGGTGCTCCGCGCGCTCTCCAGTGAAGGCGCAGGTTCTCCCTGACCGG 90965  
QY 3071 TCGAAGCGCGCACGCTTGGCGGTGATGCTGATCTGCTGTTGTTCAAGCCGTTACAG 3130  
DB 90964 CGGAAGCGCGCGGCTTGGTGTGTTGCTGTTCTGTTCTCTGTTGCTGACGCCGTTACAGC 90905  
QY 3131 GAGGTACGGAACCGGCGCGGACGATCCGACAGCGGGCTGGCGGCGGAGGTCATGTCGATG 3190  
DB 90904 GAGGTGCGGATGCGCTGACGACACAGCCCGGCTGGCGCTTCAGGTTGCGGTGATG 90845  
QY 3191 AGGGGGAAGCGTGTGTGTGAGACGGTCAAGGCGGAAGCCCGCGCGGCGGTGACAC 3250  
DB 90844 AGCGGTGAGCGTTGTGTGACGCTCATGTGAGGCGCGCGCGCGCGCGCGCGG 90785  
QY 3251 AACGCGCGGATACCGGGTGTGATCAAGTTCAAGCGCGCGGAGAGATCATGCGCGT 3310  
DB 90784 AGTTGCGCGACGCCCGGATGTGATGAGCGGTTCACGCCGCGGAGAAAGTACGCTGAG 90725  
QY 3311 GGGTCTCCGACGCGGACCTTCGTCGATGACCGGAGTGAAGATCTCGTTGCGCGGACAC 3370  
DB 90724 GGTGTGTTCGAGGACCTTCGTCGATGAGAGCGGAGCATGAGCTGTGCGCTTTGAGC 90665  
QY 3371 GACTCCGCGTCTACTTGGACCTGTCAACCCGTACGAGAAATGACAGCGGAACTGAC 3430  
DB 90664 GGGAGGTCTTCTGAGCGCGCGCGGTGAGCGCGACACAGAAATGAGAGGAACTGAC 90605  
QY 3431 GTCCGCGCGGGGTAGAGCCGACGCTGTACGAGAAAGCGGCTCCCGCGTTACCGCGAC 3490  
DB 90604 GTCCGCGCGGGATACGACCGACTGTGAGGAAAGCGCGGCGCGGTGTACCGCGGCG 90545

QY 3491 TCCAGCGCGCGCGGCTTCAAGCGCGCTGATCGTGTGCTGAGTACTTGCCTCCGCGG 3550  
DB 90544 TCGAGGCGCGCGCGCTCCGCTCAAGCGGAGGATGTGTGTTGCTCAGTACTTCCCTGCGG 90485  
QY 3551 CCTGCTTCAACCGCGGTGCTTCAAGTGTGGGATGCTGATCAACAGTGTGAGACGTGCGG 3610  
DB 90484 CCGCGTGCACGCGCGGTGCTCAGCTCGGTGATACCGGTGAAAGAGACCGAGCGCGGCG 90425  
QY 3611 AACCGGCGCGGCTCAAGGTGAAACCGAGCGGACCTGCTCAACCGGGGTGAACGCGAG 3670  
DB 90424 AAGCGCGGGGGGTTCAGCGCCAGAGGTGAGCGGCGCTTTCAGAGGTGTGAAACGGATG 90365  
QY 3671 GCGCGGTAGTGGCGGTGAGTTCGCGAGCGCGGACCGCTGTGCTTGGCGGTGTCGAT 3730  
DB 90364 GAGCGGTACCGCGGTGCGCAAGGTGACGAGGTGCTCCGCGCTGTCGCGGCGCGCTCCG 90307  
QY 3731 GCGGACCGGTGAGGCGGGTGACTCGGCGCTGAGCGCCCGCACAGGCGCGCGGTG 3790  
DB 90306 -CGGTGTGCGCGAGCTGGCGGAGGCGGATGCGACCGCTTCGATGCGCGCTGCAGCGT 90248  
QY 3791 GCGTGGGTGCTGCGCGGTCCGCGATCTCCGTGCTACGGCGGTGCGCTGATGATGATC 3850  
DB 90247 CTTCGAGGAGGTGCGCGGTCTGTATGCGGGGGCGGGCGCTGCGCTTGGACGCTG--- 90191  
QY 3851 TTCAATGACGTGCGAACCTTTCGCGCTGTGTGTCGAGAGATCAAGAACCTTTCGCTT 3910  
DB 90190 -TCCAGAGAACTGTATCTCCGCGTGCAGACTGGGAGACGAGCGCGCTCAGAGAGCG 90132  
QY 3911 CCGCTTGTCCCATCTCCGCTGTATCAACGCAACCGCTGATGTGCGAGTGGAGTGAAGTC 3970  
DB 90131 TGTGTGGCCCACTGCATGATCTGCGCGCGCTGCGCGGAGCGAGATGATGCT 90072  
QY 3971 GCACACCGCGGAGTGTCTCACTGACACCGAGGCGCGGTGGGAGGCGAGACCG 4030  
DB 90071 GCACACCGGAGATGCTTCGAAAGACACCGAGCGCGGTGGGAGCGGATCACTG 90012  
QY 4031 TTCGCGAGCCACTCGGTGTGTGCAACCGGAGCGGTGCTCGTGGCGTGAAGCGTAT 4090  
DB 4091 CTGTTGCGAGCGGGGAGAAATGAGGTGTGTGCAACGACCTTCTCCGCGCGAGATGCG 4150  
QY 89951 CAGATGCGAGGGGAGTGAAGATCAAGCTGGGAGAAACCTTTCGCTGACAGAGG 89892  
QY 4151 CTGAGCGCGTCAAGTGCATGATGCGGAGTGGCGGCTCCGACAGAGATCACTG 4210  
DB 89891 CCGAGCGGTGCGGAGTGCAGACGAGTACCGCGCGGTCTCCGAGAGTCACTG 89832  
QY 4211 GTAGTTGCTCTCTGTCGCGGAGGAGGAGTGAACGCTGACGCTGACGTGCGGAG 4270  
DB 89831 GTGGTTGTGAGTTCGTGCTGTGGAAGAGACCTCGGTGCGGGAGCGCGGTGAG 89772  
QY 4271 CTGCTGTGTGAGAGCGGTGTGTTGACCCGATGTGTCTCTGTTCTCGGAGAACGCGTC 4330  
DB 89771 TTCCGCGTGTGAGGCGGTGTTCTCAAGTTCGCTGCGGGTTCGCGGAGAGGCTTC 89712  
QY 4331 GAGGAGGTGAGCCCATGAGCGCGGAGCATGCGTCACTTTCGCTGTGTGTCGATCTC 4390  
DB 89711 GAGGAGGTGAGACCATGAGGCGCGGAGCATGCGTCACTTTCGCTGTGTGTCGATCTC 89652  
QY 4391 GGTGACCACTTGTTCGCGCGGATGCGAGTGTGATGAGCCCGGATCGTTCGCGGAG 4450  
DB 89651 GGTCAAGGACTTGTTCGCGGTGATGCGGAGTGTGAGAGGCGCACAGCCGTTGCGGAG 89592  
QY 4451 CAGGCGGTGCTGAGTGAACACCGCGCGCTTCGAGAGCGGTGATCGGCTTGTGTGCGT 4510  
DB 89591 TTCCGCGGTGCTGAGTGAACACCGCGCGCTTCGAGAGCGGTGATCGGCTTGTGTGCGT 89532  
QY 4511 GAGGTTGAACACTTCGCGGTGATCAAGGATCGCGGACCGGAGTGCACCGCTGTGACAG 4570  
DB 89531 GAGGTTGAACACTTCGCGGTGATCAAGGATCGCGGACCGGAGTGCACCGCTGTGACAG 89472  
QY 4571 CAGGCGGTGAGCGGTGAGAGAGCTTGAACCTGTGTGTCGCGGAGTCTTGCAG 4630







Db 10903 GCGGCGCGTGAAGACCTGCTGAGAGCGCGTCCATGATGACTCGGCGCGGCGCGC 10962  
Qy 2515 GATGCTGGGCGCGACGTGACGAGATGTGTGAGACGAGGTGCGAGGCTCCGGT 2574  
Db 10963 CACCGGCGCGCGGCTTGAAGTGTGCGGAGCACTCGGTGAGCGAGGTGCGGCGCC 11022  
Qy 2575 GACCGGCGCGGCGAGCTGAGCGGCTGAGCGCGCGCGAGCGCGGAGAAACCGGCTCCGGTGA 2634  
Db 11023 CACCGGCGCGCGATGATACCGGCTGCGCGCTCGAGGTCCGGAGAGCGCGCTTCGGCTGA 11082  
Qy 2635 CAGGTAGACGTCTCCGAGCAGGTGATCTGCAACCGCACTTGCGAGTGGCGGTCCGCGC 2694  
Db 11083 CAGGTACAGTCCGCGAGAGGTGACTGCAACCGCACTTGCGGAGTGGCGGTGGCGCG 11142  
Qy 2695 CATGTTTCCGCGCGGATTCGCGCAACACTGCGGCGTCCACACCCGAGCGAGCTCTCA 2754  
Db 11143 CATGTTGCGGCGCTTGAATCCGCGAGAGTTCGGCGTCCGCGCGCTTCAGGCGTTCAG 11202  
Qy 2755 CGCGTAAACCGAGTCCGCTGCAATGCGCGGAGTCCGCTCCGCGCGGTAGTCAAGAAC 2814  
Db 11203 GCGGTAGCTGATGCTGATGAGACCCGCGGCTCCGCTCCGAGACCCGCTCTCGAAGGC 11262  
Qy 2815 GCGGAGGCGCTCCGCGCACTCCGCTCCGCTCCGCACTTCGCGAGTCCGCTCCGAGC 2874  
Db 11263 GCGGAGGCGCTCCGCGAGCTCCGCGCGCTCCGCTCCGCGAGAGCTTCGCGCGCC 11322  
Qy 2875 GCTGTAGTCTTCGCGAGCGGTGACGAGTGAAGCGCGCGGTGCGGCTGAGCTCTGTTGAG 2934  
Db 11323 GCTGTAGTCTTCGCGAGTGTGAGAGAGTGTGATGATGATGATGATGATGATGATGATG 11382  
Qy 2935 TTGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2994  
Db 11383 GTGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11442  
Qy 2995 GTTGAAGCGCGCGAGCGCGAGCGCGCTTCGCGCGCGATCCGCGAGAGTCCCGAG 3054  
Db 11443 GCGGAGCGCGAGGTGATGATGCGCGCTTCGCGCGCGAGAGTCCGCGAGAGCGCGCAG 11502  
Qy 3055 GTTCTTCTGATCCCGTTCGAGAGCGCGAGCGCTTCGCGGTGAGCGCTGATGATGATG 3114  
Db 11503 GTTCTTCTGATCCCGTTCGAGAGCGCGAGCGCTTCGCGGTGAGCGCTGATGATGATG 11562  
Qy 3115 GTTCAAGCGCGTACAGGAGGTGAGAGCGCGCGAGGTCCACAGCGCGCGGTGAGCGCG 3174  
Db 11563 GTTGAAGCGCTGATGAGCGAGGTGAGAGCGCGGTGAGAGCGCGCGAGTCCGCGAGT 11622  
Qy 3175 CAGCGTCTGTTCCGCTGAGAGCGCGAGCGTGTGTGAGAGCGGTGAGAGCGCGCGCG 3234  
Db 11623 CAGCGTCCGCTCCGCTGAGAGCGCGAGAGTGTGTGAGAGCGGTGAGAGCGCGCGCG 11682  
Qy 3235 GCGCGCGTGGAGACCAACCGCGCGATCCCGGAGTGTGATGAGCGGTTCAGGCGCGCG 3294  
Db 11683 GCGCGCGTGGAGCGCGAGCGGTGAGAGCGCGGAGTGTGAGAGCGGTTCAGGCGCGCG 11742  
Qy 3295 GAGGTACATGCGCGGTGAGTGTTCGAGCGCGACCTGTGTGATGAGCGAGGTGAGAGT 3354  
Db 11743 GAGGTACAGCGCGGTGAGTGTTCGCGCGGTGATGATGATGATGATGATGATGATGAT 11802  
Qy 3355 GTTTCGCGCGACCAACCGCTCCGCTGATCTTGTGACCTGTGACCTGATCCGAGAGT 3414  
Db 11803 GTTTCGCGGTGTGAGAGCGCGAGCGGTGTGAGCGCGCGCGGTGACCGAGAGCGAGAG 11862  
Qy 3415 ACAGCGAGACATGAGTGTGCGCGCGAGGTGAGAGCGCGAGCGGTGATGAGAGCGAG 3474  
Db 11863 GCGCGCGAGACATGAGTGTGCGCGCGAGGTGAGAGCGCGAGCGGTGATGAGAGCGAG 11922  
Qy 3475 CCGCGGTGACCGAGCGCTCAACCGCGCGCGGTGAGAGCGCGGTGATGAGAGCGAG 3534  
Db 11923 GCGCGAGAGCGCGAGGTGAGAGCTCCGCGCGGTGAGAGCGAGCGAGGTGATGAGAG 11982  
Qy 3535 GTACTTGTCCGCGCGCGCTGATCAACCGCGCGCGGTGAGAGCGAGGTGATGAGAG 3594

Db 11983 GTAGCGCGCGCGCGCGCGCTTCAACCGCGCGCGAGGTCCGCGAGACCCCGGAGACAG 12042  
Qy 3595 GTGAGAGAGTCCCGCGAACCCCGCGCGGTCCAGGTGAGACCGAGCGGAGTGTCTCA 3654  
Db 12043 GCGAGAGAGCGCGAGAGCGTCCCGGTGAGCGCGCGAGGTGAGAGCGGTCTTCA 12102  
Qy 3655 CCGAGTAAACCGAGAGCGCGGTAGTGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 3714  
Db 12103 CCGAGTAAACCGAGAGCGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 12162  
Qy 3715 TTGCGCGGTGTGAGTGTGAGCGCACCGGTGAGCGGTGAGCGGTGAGCGGTGAGCG 3774  
Db 12163 TTGCGCGGTGTGAGCGCACCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCG 12222  
Qy 3775 CAGCGCGCGCGGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCG 3834  
Db 12223 CCGCGCGCGCGGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCGTCCGCG 12282  
Qy 3835 CGCGTAAAGAGTGTCTTCAATGAGCGAGACCTTCTGCGGTGAGTGTGAGAGATC 3894  
Db 12283 GCGCGGCGCGGTATGAGAGCGTCCATGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 12342  
Qy 3895 ACAGACCGGTGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 3954  
Db 12343 GCGAGCGCGCGCGGTTCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 12397  
Qy 3955 CAGTGTGATGAGTGTGAGCACCGCGCGAGTGTCTCACTGAGCACCGAGAGCGGTG 4014  
Db 12398 AGAGTCTCAAGAGTGTGAGCACCGCGCGAGTGTCTCACTGAGCACCGAGAGCGGTG 12457  
Qy 4015 GAGAGCGAGCACCGGTGAGAGCGCACCTGCGGTGAGTGTGAGCACCGAGAGCGGTGAG 4074  
Db 12458 GAGAGCGAGCACCGGTGAGAGCGCACCTGCGGTGAGTGTGAGCACCGAGAGCGGTGAG 12517  
Qy 4075 TTGCGTAAAGCGTCACTGAGTGTGAGCGCGAGAGTGTGAGCGGTGAGCGGTGAGCG 4134  
Db 12518 GCGGTGAGCGGTCACTGAGTGTGAGCACCGCGAGAGTGTGAGCGGTGAGCGGTGAG 12577  
Qy 4135 CCGCGCGAGAGTGTGAGTGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAGCGGTGAG 4194  
Db 12578 CCGCGTTCAGAGCTTCACTCAACAGAGTGTGAGCGGTGAGCGGTGAGCGGTGAGCG 12637  
Qy 4195 GAGTATCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4254  
Db 12638 CAGAGTCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 12697  
Qy 4255 CCGGTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 4314  
Db 12698 CCGGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 12757  
Qy 4315 TTGCGAGAACCGGTGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4374  
Db 12758 CCGTGTGAGAACCGGTGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 12817  
Qy 4375 CCGTGTGAGAACCGGTGAGAACCGGTGAGAACCGGTGAGAACCGGTGAGAACCGGTGAG 4434  
Db 12818 CCGTGTGAGAACCGGTGAGAACCGGTGAGAACCGGTGAGAACCGGTGAGAACCGGTGAG 12877  
Qy 4435 GAGTCCGTTCCGCGAGCACCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4494  
Db 12878 GAGTCCGCGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 12937  
Qy 4495 CCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4554  
Db 12938 CCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 12997  
Qy 4555 CACCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4614  
Db 12998 CCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 13057  
Qy 4615 CCGCGATCTTGTGAGCGCGTTCACAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 4674  
Db 13058 CCGCGATCTTGTGAGCGCGTTCACAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 13117

Qy 4675 TCAGCGCGGCTCTGCGGGGATGACGAGCGGCGACGTGATCGAGGTCGACGACCGGTCG 4734  
Db 13118 TGGCCGAGGTCGCGGGGCGTGAACGCGCGCGCCACCGCGTCCGGGTCCAGGGTGCCTGT 13177  
Qy 4735 CCGGGTCGACGTGCGAGAACACCGGTGTGATCCGAGCCAGCTGCGCCGCTGCGCGGTG 4794  
Db 13178 CCGGGTCGATGTCGCGCAAGACCGGGGTGAGCGCGATCCAGCGCATGCGCTGCGGGGTGG 13237  
Qy 4795 CCGGGAAGTCATGACGCGCATGATCACTTCACCGGTGAGCGTCAACCGGCGCGGACCA 4854  
Db 13238 CCGGGAAGTCATGACGCGCATGATCACTTCACCGGTGAGCGTCAACCGGCGCGGACCA 13297  
Qy 4855 GTTCAGAGCGCGACCGGTGCGCTTGGAGGTGCGATGATGACCGGTGACCGGACGAGTCG 4914  
Db 13298 GCTGGAACCGCGCGCGTGGCGTTCGAGGTGCGCAACGCGTGTTCACCGCGCGGACGCG 13357  
Qy 4915 CGACCCGCGGCTCGAATCTCCGCAACGAGGGCGCGCGTGTGTGACCGCATGTTGTTCA 4974  
Db 13358 CGACCGGCTCTCTGAACTCGCGGACGAGCGGGCGCGTGTGACGACCACTGCGCGTCA 13417  
Qy 4975 GGGCCCACTCGAGCGCGGCGGACGAGACCGCTGCGCGTCCGCGGACGATGCGCGCGGCG 5034  
Db 13418 GCGCGCGGTGAGCGCGCTCTGAGAGCTTGCGCGCGGTGATGCGGTGCGCGCGGCG 13477  
Qy 5035 GCAGCGGGGTGAGAGCGCGCTCGGAGACCGCGGACGAGATGCGGACGATGCGGTGCGGCT 5094  
Db 13478 GGAACGCGGTGATCGAAGCGCGCGGTGCGCGGCGGAGAGATGCGAGGTGCGATGACGCTTT 13537  
Qy 5095 TCACG 5099  
Db 13538 TCACG 13542

RESULT 11  
AF079762 12441 bp DNA linear BCT 27-OCT-1998  
LOCUS AF079762  
DEFINITION Streptomyces venezuelae desosamine biosynthetic gene cluster,  
complete sequence.  
ACCESSION AF079762  
VERSION AF079762.1 GI:3789892  
KEYWORDS  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 12441)  
AUTHORS Xue, Y., Zhao, L., Liu, H. W. and Sherman, D. H.  
TITLE A gene cluster for macroide antibiotic biosynthesis in  
Streptomyces venezuelae: architecture of metabolic diversity  
Proc Natl. Acad. Sci. U.S.A. 95 (21), 12111-12116 (1998)  
JOURNALS 9770448  
PUBMED 2 (bases 1 to 12441)  
REFERENCE 2  
AUTHORS Zhao, L., Liu, H. W. and Sherman, D. H.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUL-1998) Department of Microbiology, University of  
Minnesota, 420 Delaware Street SE # 1060, Minneapolis, MN 55455,  
USA

FEATURES  
SOURCE  
1. 12441  
/organism="Streptomyces venezuelae"  
/mol\_type="Genomic DNA"  
/strain="ATCC15439"  
/db\_xref="ATCC:15439"  
/db\_xref="taxon:54571"  
119. 1327  
/gene="desVII"  
119. 1327  
/gene="desVIII"  
119. 1327  
/gene="desVIII"  
/codon\_start=1  
/transl\_table=1  
/product="putative tautomerase"  
/protein\_id="AAC68676.1"

/db\_xref="GI:3789893"  
/translation="MTDITLGTALTOPPLGRTVRAVADRELGLHLETRGIIHWIHAANG  
DPATVLRGQADDPDYAEKVRARALSFSTGSWTADALAASTLCSYDFVSGAD  
GVPPQVQVSGGCGLEERQVPAAGDVGQORAVRAGIHEHLEGLAPDSASYA  
FELLGFPVAPATVAAAAYGVADRADPADLLERLRPLGSDLSLAPQSRVRAAG  
ALMELTALDSDSPGALLSALGVTAAVDTLGNATALLAHPEQRELCDRPDLAA  
AVEETLRYPDPVOLDAVRGTELRGLRPAQAHVLTATGDRDEVTDEPRPL  
ARPDAAHNLHPAGVPAVSLVRLQAEVALTLTLAGFGLRQAGDVLRPRAPVR  
GPLSVPSGS"  
1475. 2755  
/gene="desVII"  
1475. 2755  
/gene="desVII"  
/codon\_start=1  
/transl\_table=1  
/product="glycosyl transferase"  
/protein\_id="AAC68677.1"  
/db\_xref="GI:3789894"  
/translation="MRVLTSTFAHHTTYGLVPLANALLAAGHEVRAVAGPALDTIT  
GSGIAVPVGTDLIHEYRVMAQEPNPNPALAFDARPEPLDMHAGIETALAY  
FHLIANNDSWVDLVDFAVSMQDVLVMEPTTAVAGAAQVTAHARVWGPVMS  
ARRKFVLRDRQPERHEDPTAEMLTTLDRYGASFEEELTQFTIDPTPSRLDT  
GLPTGMRVYPNGTSTVPMLSBPAPRVCULTIGVARSVIGDVGSDILLEALA  
DLIEVATIDAGRAIRIRYPKTRTPTDVPMAHLLPSCSAILHGGAGCTAAVLA  
AVQVMLAEIAMDAPVKARAAVQAGFFLPALDTPOAVRDAVRLIDPSVATAAR  
LRSTFQDPTPAGIVPELRLAQRPPADAR"  
2845. 3558  
/gene="desVI"  
2845. 3558  
/gene="desVI"  
/codon\_start=1  
/transl\_table=1  
/product="N,N-dimethyltransferase"  
/protein\_id="AAC68678.1"  
/db\_xref="GI:3789895"  
/translation="MYEVDADVDYDLYLRGQDYAAASDIADLVSRPPEASLID  
VACGTGHLHFTEFEDTAGLEISDMLTAKRLPDATLHQDMKDFRIGKFSAY  
VSMPSVGYKTTTEELGAASFAEHLDPGVVVEPWPPEPTADQVMSADVVRDG  
RTVARVSHSVREGNATMEVHFTVADPGKVRPSVHLITLTHQAEYEAFAAAGLR  
VEYLEGSGRGLFVGVA"  
3625. 6054  
/gene="desR"  
3625. 6054  
/gene="desR"  
/note="putative extracellular protein involved in a  
resistance mechanism"  
/codon\_start=1  
/transl\_table=1  
/product="beta-glucosidase"  
/protein\_id="AAC68679.1"  
/db\_xref="GI:3789896"  
/translation="MTGKTRILPRVRKRTTPRAFLAVVGLTAGTTVAAPAAADT  
ANQYTSRAAEVVAQTLDEKISFVMAHALDPRDNGVLTGPVPLGIPELSAAGPNG  
IRLVGQRTALPAPVALASTFDTDMAISYGMKRDGRALNDQVLTGPMNNITVPHG  
GRNTEFSBDPLVSSRTAAVQIKQIQAQGLMTAKHPANNQENRFSVNAVSDQL  
REIEFPAPFASSKAGASFMCAVNGLKGSCGNDLNNVLRQVQGVMSDMLA  
TPGDDATITKGLDQMGVELPDGVKGPSPPAKFEGALKTAVLNGVPEAAVRSAS  
RIVGOMEKGLILATPA PRPERDKAGAAVSRKVAENGAVILRREGQALPLAAGKGS  
IAYGPTAPVDPKVTGSGSAHYVPSAAPLDTITAPRAGAATVYVGESEFTGQDIPA  
GNLSPATNQGHLSPGKAGALYDGTITVPADGERIARVATRGYATVQDGHSTTEAGQ  
VYGVSSPLKHLITGTHKLTISGFAMASVPLISLGWTPPAADATTAKAVASAKAR  
TAVFAVDDGTEGVDRPMLSLPTQDYLISAVDANPNTIVLVTGSSVLMPLMSKTR  
AIVLMMYDQAGAEATAPALAYGVNPSKGLQSFPALENQAVAGDPTSPYGVNQQT  
YREGIHYGRMPDKENYKPLFPFGHGLSYTSFTQSAIPVTVTSGGLKTVTVNSGK  
RAGGVQAVIAGASPNTYAPAKKULVGYTYSVLSAAGAKTVTVNVDKRLQTSBSA  
DLRGSATVNV"  
complement(6143. 7282)  
/gene="desV"  
complement(6143. 7282)  
/gene="desV"  
/codon\_start=1  
/transl\_table=1  
/product="transaminase"

gene  
CDS

gene  
CDS

gene  
CDS

gene cds  
/protein\_id="AAC68680.1"  
/db\_xref="GI:3789897"  
/translation="MSSRAETPRVFLDLKAAVEBELAETDAIARVLDGSGVTLGPE  
LEGPRFAAYCETDHAIVNGMDLQIALRGIGRGDEIVPSTRTYASLTAVA  
TGAIPVPERHEHEDPTDPLVKEATIRPRALPHLYGHPRADMRLRELADRGH  
IVEDAQAQHGAKTRGRIAGSSVAATSPFGKNLGFQGGAVTGDPELARIKML  
RNGSRQKSHETKINSRLDEMOAVLRRLAHLDWNGRRSLAAEYSLGAGLPG  
IGLPVAPDTPDPMVHLPTVTERDERLSHLDAAGIDTTLTHYPPVPLSPAYGEAP  
EGSLPRASFARQVLSLPIGPHLERPOLAVITDAVREMARVDA"  
/complement(7279..8292)  
/gene="desiv"  
/complement(7279..8292)  
/gene="desiv"  
/codon\_start=1  
/trnasl\_table=1  
/product="TDP-glucose-4,6-dehydratase"  
/protein\_id="AAC68681.1"  
/db\_xref="GI:3789898"  
/translation="MRLVTGAGFISGSHFROLLAGAPDPADEVITVLDSTVAGN  
RANLAPVDADPRLEFVGDIRDAGLARELGVDAIVHFAESHSVDSIAGSVFET  
NVOGTOTLQCAVDAGVGVVSTVDVSDSISGWTSSPLEPNSPVASAGSDL  
VARAHRTVGLDVIRCCNNVGYOHPGLIPLPYTNLIDGGLPLPYGNAVREY  
HTDHCKGLMLVLAGGAGEIYHIGGLLITNRBLTGILLDSGADMSVRAADRKG  
HDLRYSIDGGEIERELGVRPQVSPADGLARTVWYRENGMWEPLKATAPLATAVE  
VSA"  
/complement(8424..9302)  
/gene="desiit"  
/complement(8424..9302)  
/gene="desiit"  
/codon\_start=1  
/trnasl\_table=1  
/product="alpha-D-glucose-1-phosphate  
thymidyltransferase"  
/protein\_id="AAC68682.1"  
/db\_xref="GI:3789899"  
/translation="MKGIYLAGSGGTRLHPATSVISKQILPVYKEMTYLPSVLMIG  
GIREIQIISTPHIELFOSILANGRHGIELDYAVOGEPAGIADALLVGAHIGDTC  
ALLIGNIFHGPGLYTLRLDSIARLDGCVLFGVPDPERGVAVDAITRLDIYERK  
PVKPRSNIAVTGLYLDNDVVDYAKNIRPSRGELEITDYNRYILRGRAELVNLRG  
FAMLDGTGHSLSLRMAQYOVVLEERGCVMLGLEIAFRMGFLDAERCHGLGSLRT  
EYGSYLMEIAGRGAAP"  
/complement(9495..10952)  
/gene="desiit"  
/complement(9495..10952)  
/gene="desiit"  
/codon\_start=1  
/trnasl\_table=1  
/product="putative reductase"  
/protein\_id="AAC68683.1"  
/db\_xref="GI:3789900"  
/translation="MTRAPLSATAPARCAHPADLGAHVAVGOTLAAGLVPRDEA  
GTTARLVRILAVYGNSPPTPLEARHDLGVDAIRRLALFGVPELRTAVETPA  
GAWKNITLPLEORGVDALARKPVPYSGVLYPPTCMARFPCVRVYGAIVTSPA  
LAGANAFRSVIDEIPAGNPSAMYFSGGLEPLTNPGSLAAATDHGRPTVYTSF  
ALTERLIEROPGLMGALIRTSLYGNDYEBQTTKKAFRRVRENLRFPOLRER  
ESPINGFAYIVLPGASRLDLVDPIADLNDAGQCRKTDPVIRREDYGRDGRKLPQ  
ERRASIOBALNAFERVRETTPELHIDYGTALNSLRIGADAEILRIKPYMRPTAPRO  
VAOVLDLGVYLYREAGFPDLDGATRYLAGRVTPDTSITLEVVRDEVERGEVAANDG  
DEYFMDGPDQVYARLNOLEERDADGMEARGLR"  
/complement(10949..12196)  
/gene="desiit"  
/complement(10949..12196)  
/gene="desiit"  
/codon\_start=1  
/trnasl\_table=1  
/product="4-dehydroase"  
/protein\_id="AAC68684.1"  
/db\_xref="GI:3789901"  
/translation="MKSALSDLAFFGGRPAFDDPLVGRPNRIDRARTYERLDRLDLS  
QWLSNGPLVREFEERVAGLAGVRAVATNATNATAGQLAAAGLGEVIMSMTPAA  
TPHALMIGLTPVADIDPDTGNLDPDVAAATPRTSLAVGVLHMGRCADQDLKX  
ADBEHGLRYFDAHALGACAVDRPAGSLDAEYFSHATKAVVAFEGGAVDDADLA  
ARIRALHNGFDLPGSGPAGTNAKMSAAAMGLSLDAFPEVIDRNRNNAAYREH

ORIGIN  
Query Match 14.6%; Score 1309.4; DB 1; Length 12441;  
Best Local Similarity 67.7%; Pred. No. 6.3e-05;  
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;  
LADLPGVLVADHDHRLNHNQYIVIEIDATYIHRDLVMEVLKAEGRVTRAYFSPGC  
HELEPYRGOPHAPLPHTERLAAVSLPSTALGDDIRRVADLRLCARHSELTRAR  
HDTAPAPLAAPOTSIPTIGRSR"  
QY 2369 GGGCGGATCGACCCGTTGACAGAGCGCGGATCAGGCGCCGATGCTGATCCACCGTCG 2428  
DB 9472 GGGGGGATCGGGGCGGGTGGGTGCGCGCAGAGAAAGCGGGCGCTCCCGACCGCTCC 9531  
QY 2429 GCGACGTCCTCGTTCGAGTTGGTTAGCGCGGGCGGTTCACCGACTGTCGAAACCGTGAAG 2488  
DB 9532 GCGCGCTCGCGCTTCAGCTGCTTCAGCGGGCGGTTCAGCACTGATCGAAGCGCTGCATG 9591  
QY 2489 AAGAACTGTCCTCCCGCGCTGCGATGATGCTGCGGCCCGACGTCAGAAAGTGTGACG 2548  
DB 9592 AAGTACTGTCGCTCGCTGACGCGCGCCACTCGCCCGCGCTTCGAGAAAGTCCGTGACG 9651  
QY 2549 ACCGAGTGAAGCTCGGTCGCGGTGACCCGCGCGAGCGTAGCGGGTGCGCCCGCC 2608  
DB 9652 ACTTCGCTGAGGAGGAGTGTGCGGGGTCAGCGCGCGCGAGTGAAGCGGGTCCGCTCC 9711  
QY 2609 AGCCCGGGGAAACCGCGCTCCCGGTACAGGTAGACGTCTCCGACAGGTGATCTGAC 2668  
DB 9712 AGGTGAGGAAAGCGGCGCTCGCGGTACAGGTACAGTCCGCGAGATGACCTGAC 9771  
QY 2669 GCCACCTGCGATGAGCGGTCGCGCGCATGTTTCGCGCGAGTCCGCAACAGCTGGCG 2728  
DB 9772 GCGACCTGCGGCGGTGCGCGGTGGCGCGCATGTGCGCGGCTTGAATCCGACAGTTCCGCG 9831  
QY 2729 TCGACACCCCGACGACGCTCTTCAACGCGTAAACGAGTGTGATGCGCGGCTC 2788  
DB 9832 TCGCGCCCGGTGCGACGCTGTTCAAGGCGGTAGCGTGAAGTGAAGTCCGCGGCTG 9891  
QY 2789 CGCTCGCGCGGTGACGTGACGAAACCGGCGAGGCGCTCCCGCACTCGCTGCGCTCC 2848  
DB 9892 CGCTCGCGGACCGGCTCTTCAAGGCGGTGAAGGCGCTCTGAGGCTCGCGCGCTCC 2901  
QY 2849 TCGGACACCTGCGCGGTGCTCCGACCGCTGAGTCTCGCGCGAGCGGTGACGAGTGCAGC 2908  
DB 9952 TCGGACGCTGCGGTGCTGCTGACGCGCGCTGAGTCTTCCGCAATGTTACAGAGTGCATC 10011  
QY 2909 GCGCGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2968  
DB 10012 GTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10071  
QY 2969 TTGCGCGCGCGCGCGCGAGTATGTGTGAGGCGAGCGCGACCGGTGCGCGCTGCG 3028  
DB 10072 GAGGCAACGCGCGCGCGCGAGTATGTGTGAGGCGAGCGCGAGTATGCGCGACCTGCG 10131  
QY 3029 GCGCGCATCGCGAGAGGCGCGGAGGCTTCTTCTGACCGCTTCAAGGCGCGACGCTGCG 3088  
DB 10132 GCGCGCATCGCTGAGAGGCGCGAGTCTGCGCGACCGCGCGAGGCGCGCTTCTTCTG 10191  
QY 3089 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3148  
DB 10192 CCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10251  
QY 3149 AGGTCCCAACAGCGCGCGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3208  
DB 10252 AGGCGCCCAAGAGCGCGCGCGCGCGCTTCCAGGCTGCGCTCGGTGAGCGCGAAGGAGTTGCTG 10311  
QY 3209 TGAAGGTGAGGCGAGCG 3268  
DB 10312 TGAAGGTGAGGCGAGCG 10371  
QY 3269 TTGCTCAACGCTTTCAGCGCGCGCGCGAGAGTCACTGCGCGCTGCTTTCGACGCGAC 3328  
DB 10372 TTGCTGAGCGGCTCCAGCGCGCGCGCGAGAGTCACTGCGCGAGAGGCTTTCGCGCTATC 10431



QY 3329 TCGTGTATGACCGAGGTGAGAGATCTGTTGCGGAGCACACCGATCCCGCTGCTGATCTTG 3388  
Db 10432 TCGTGTATGACCGAGGAAATAGCGCTTGGCGGCTGAGAGGCGAGCGGCTGTAAGCGG 10491  
QY 3389 GCACCTGTGACCCCTTACGACAGAGTGAACAGGAGAACTATGCACTGTGCGGCGCGGGTTAAGC 3448  
Db 10492 GCGCGGTGACACGAGACGCAAAAGTGGACAGCAATGACAGGTGCGGCGCGGGTTAAGG 10551  
QY 3449 CCGACGCTGTACGAGGAAAGCGGGCTCCCGGCTGACCGGAGCTTCCAAACGCGCGGCGGCGG 3508  
Db 10552 CCGACGCTGTACGAGGAAAGCGGGCTTCTTGAGCGAGCGCGCTGTGAAAGACGCGCGCTGT 10611  
QY 3509 TTACGCGGCTGATCTGTTGCTCTCAATCTTTCGCGCGGAGCCCTGTCTCACCGCGTG 3568  
Db 10612 TCGAGCGGAGACAGAGGTGTTCTTCCAGATACGCCCGGCGGGGCGGCTGTCTGACCGCGTG 10671  
QY 3569 CCGAGCTGCGGAGATGCGATGCAACAGTTCAGACGCTGCGGAAACCGCGGCGCGGTCCAGG 3628  
Db 10672 CCGAGCTGCGGAGCTTGCCTCCAGACAGGCGCGAGAGGCGCGCGAAAGCGTCCCGGTTCAGC 10731  
QY 3629 TCGAACGAGCGGCGCATCTGCTCCACGAGGGGTGAACCGGCGGAGCGCGCTAGTGGGCGGTG 3688  
Db 10732 CCGAGCTGCTGAGCGGCGCTCTCTTCCAGCGAGGTGAAGGGGCTGTTGCTGCTGACGACGCG 10791  
QY 3689 AGTTCGAGGAGCGGAGCCGCTGTCTGTGCGCGGCTGAGATGCGGACCGGTGAGGCGG 3748  
Db 10792 AGCGGAGACGAGGTGGCGGGCGGCTGTCTGCGCTGTGCGGCGGAGCGCGCGCGCG 10851  
QY 3749 GTGACCTGCGGCTGAGAGCGCGCGCACCGGCGCGGCTGCGGCTGCGGCTGCGGCGG 3808  
Db 10852 GCGAGGCTGTGCGCGAGCGGCTGTGACCGCGCGCCCAATGCGCTCCCGGGTGTGCGCGAG 10911  
QY 3809 TCCGCGATCTCGTGTGATGAGGCGGCTGCGGAGAGTGTCTTCAATGACGTGGAAC 3868  
Db 10912 CGTTTCGCGCGGCGGCGGTGAGCGGAAAGCGGCGGCGGTATCGGAGGCTCAATCGG 10971  
QY 3869 CTTTCGCGGTGTGTGTGAGGAGATGACGAACCGTTGCTTGTGCTGCTCCACTCGCG 3928  
Db 10972 GCGTGAATGTCTGCGGCGGCGCGGAGCGGCGGCGGCTGTGCGGCTGCGGCGGCGG 11026  
QY 3929 GTTGTATCAAGCAGCAGCTGT 3988  
Db 11027 GGTCAATTCGCGCGCGGCGGT 11086  
QY 3989 CTCACTGAGACCGAGGCGGCGGT 4048  
Db 11087 GTCTGTGCGGAGTGGGCTGT 11146  
QY 4049 CTGTGTCAAGCGGAGCGGCTGT 4108  
Db 11147 GTGCGGAGCGGCGGCGGT 11206  
QY 4109 GAAATGAGGCTGT 4168  
Db 11207 GAAATGAGGCGGCGGT 11266  
QY 4169 GATGCGGCTGT 4228  
Db 11267 GATGCGGCTGT 11326  
QY 4229 GAGCGGAGGAGT 4288  
Db 11327 GCGGT 11386  
QY 4289 GTGTGTGACCGGCTGT 4348  
Db 11387 GTGTGTGCGCGGCTTCCGT 11446  
QY 4349 GCGCGCGGCGGAGCTGT 4408  
Db 11447 GCGCGCGGCGGCTGT 11506  
QY 4409 GCGGATGCGGAGT 4468

Db 11507 GTGAGGCCGCAAGT 11566  
QY 4469 CACCGCGCGCGCTTCCGAGGCGGAGT 4528  
Db 11567 GACGCGCGCGCTTCCGAGGCGGAGT 11626  
QY 4529 GTGACCGGATCCGCGGAGCGGAGT 4588  
Db 11627 GTGCGCGAGGCTGT 11686  
QY 4589 GAAAGAGAGCTTGAAGT 4648  
Db 11687 GAAATGACGCGCGAGGCGGCTGT 11746  
QY 4649 TCGGCGCGGAGT 4708  
Db 11747 GCGGCGCGGAGT 11806  
QY 4709 GTGTGTGCGGAGT 4768  
Db 11807 CTGTGTGCGGAGT 11866  
QY 4769 GAGCGAGCTGT 4828  
Db 11867 GATCAGCGGAGT 11926  
QY 4829 GTGTGTGCAACCGGCGCGGAGCACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4888  
Db 11927 GTGTGTGCGGAGGCGGAGT 11986  
QY 4889 GCACTGTGCAACCGGAGCACAGT 4948  
Db 11987 GCGATGTGCGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 12046  
QY 4949 GCGGT 5008  
Db 12047 GCGGT 12106  
QY 5009 GTGCGCGATGT 5068  
Db 12107 GTGCGATGT 12166  
QY 5069 GATGCGGAGT 5099  
Db 12167 GATGCGGAGT 12197  
  
RESULT 12  
BD217172 12441 bp DNA linear PAT 17-JUL-2003  
LOCUS BD217172  
DEFINITION DNA encoding methymycin and pikromycin.  
ACCESSION BD217172.1 GI:33026942  
VERSION BD217172.1  
KEYWORDS JP 2002536959-A/2.  
SOURCE Streptomyces venezuelae  
ORGANISM Streptomyces venezuelae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
  
REFERENCE  
1 (bases 1 to 12441)  
SHERMAN, D.H., LIU, H.W., XUE, Y. and ZHAO, L.  
DNA encoding methymycin and pikromycin  
Patent: JP 2002536959-A 2 05-NOV-2002;  
REGENTS OF THE UNIVERSITY OF MINNESOTA  
OS Streptomyces venezuelae  
PN JP 2002536959-A/2  
PD 05-NOV-2002  
PF 25-JUN-1999 JP 2000557373  
PR 26-JUN-1998 US 09/105537  
PI DAVID H. SHERMAN, HUNG MEN LIU, YONGQUAN XUE, LISHAN ZHAO PC  
CI2N15/09, CI2N1/15, CI2N1/19, CI2N1/21, CI2N5/10, CI2N9/00, CI2P17/ PC  
42, CI2P19/62,  
PC CI2N15/00, CI2N5/00

CC DNA encoding methymycin and pikromycin  
FH Key Location/Qualifiers  
FT source 1..12441  
PT /organism='Streptomyces venezuelae'  
Location/Qualifiers 1..12441  
/organism='Streptomyces venezuelae'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:54571'

Query Match 14.6%; Score 1309.4; DB 6; Length 12441;  
Best Local Similarity 67.7%; Pred. No. 6.3e-85;  
Matches 1850; Conservative 0; Mismatches 876; Indels 5; Gaps 1;

QY 2369 GGGGGGGTGCAGCCCGCTTCAGAGCGGGCGGATCAGGCGCCGAGTCTCATCCAGCCGTCG 2428  
DB 9472 GGGGGGATCGGGGCGGGGTGGGGTCAAGCCGAGAGAGCGGGCGCTTCCTCCAGCCGTC 9531  
QY 2429 GCGACGTCCTCGTTCAAGTTGGTTGAGCCGGCGGTCACGCACTGGTCAAAACCGTCGAG 2488  
DB 9532 GCGGCGTTCGCTCCAGCTGGTTCAAGCGGGCGGTGACGACTGATGAAAGCCGTCATG 9591  
QY 2489 AAGAACTCTGCTCCCGCGATCGAGTCGATCGGCGCCGACGTCGAAAGTCGTGACG 2548  
DB 9592 AAGTACTCGTCGCTCGACGCGCCGCACTCGCGCGCGCTCGACGAAAGTCCCTGACG 9651  
QY 2549 ACCGAGTCAGGCTCGGATCCGGATCCGGGTCGACCCGCGCGCACTGAGCGGGTCCCGCC 2608  
DB 9652 ACTTCGATGAGAGGATGTCGGGGGTCAACGCGCCGCGATGATGAGCGGGTCCGCTCC 9711  
QY 2609 AGCCCGGGGAAACCGGCTTCGAGTACAGGTACGCTTCGAGACGATCGATCTGCAAC 2668  
DB 9712 AGGTCCGGGAGAGCCGCTTCGAGTACAGGTACGCTTCGAGAGATGACCTGCAAC 9771  
QY 2669 GCGACCTGCGGATGAGCGGTCGAGCGCATGCTTCGCGCGGATCCGCAACAGCTGAGCG 2728  
DB 9772 GCGACCTGCGGATGAGCGGTCGAGCGCATGCTTCGCGCGGATCCGCAACAGCTGAGCG 9831  
QY 2729 TCAGAACCCCGAGCGAGGCTTCGAACGCTAACCCAGGTGGTTCGATCCCGGAGTC 2788  
DB 9832 TCAGAACCCCGAGCGAGGCTTCGAACGCTAACCCAGGTGGTTCGATCCCGGAGTC 9891  
QY 2789 CGCTCGGCGGATGATGACGAGACCGGGGAGGCGCTCCGCACTCGATCCGCTCCG 2848  
DB 9892 CGCTCGGCGGATGATGACGAGACCGGGGAGGCGCTCCGCACTCGATCCGCTCCG 9951  
QY 2849 TCAGAACCCCGAGCGGCTTCGAACGCTAACCCAGGTGGTTCGATCCCGGAGTC 2908  
DB 9952 TCAGAACCCCGAGCGGCTTCGAACGCTAACCCAGGTGGTTCGATCCCGGAGTC 10011  
QY 2909 GCGCGGTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2968  
DB 10012 GTCCTGCTCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10071  
QY 2969 TTGGGCGGCGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3028  
DB 10072 GAGGCGAGGCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10131  
QY 3029 GCGCGATCCGAGAGATCCCGAGAGTCTTCCTGATCCGTTGAGAGCGCGCACTG 3088  
DB 10132 GCGCGAGCTGCTGAGAGGCGGAGGCTTCGCGGAGCGCGGAGAGCGCGCTTCG 10191  
QY 3089 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3148  
DB 10192 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10251  
QY 3149 AGGTCCCAAGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3208  
DB 10252 AGGTCCCAAGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10311  
QY 3209 TAGACGATGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3268

DB 10312 TAGACGATGAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10371  
QY 3269 TTGATCAAGGCTTCAGGCGCGCGGAGGATGATGCGGTCGGGTTCTCCAGCGCAC 3328  
DB 10372 TTGATCAAGGCTTCAGGCGCGCGGAGGATGATGCGGTCGGGTTCTCCAGCGCAC 10431  
QY 3329 TCGTCAATGACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3388  
DB 10432 TCGTCAATGACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10491  
QY 3389 GCACTGTCACCGGATGACGAGATGACGAGGATGACGAGGATGACGAGGATGACGAGGATG 3448  
DB 10492 GCGCGGTCACAGAGGATGACGAGGATGACGAGGATGACGAGGATGACGAGGATGACGAGG 10551  
QY 3449 CCGAGCTGATCGGAGAGGCGGCTCCCGCGATCGGACGCTTCACGCGCGCGCGCG 3508  
DB 10552 CCGAGCTGATCGGAGAGGCGGCTCCCGCGATCGGACGCTTCACGCGCGCGCGCG 10611  
QY 3509 TTGAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3568  
DB 10612 TTGAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10671  
QY 3569 CCGAGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3628  
DB 10672 CCGAGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10731  
QY 3629 TCAGAACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3688  
DB 10732 TCAGAACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10791  
QY 3689 AGTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3748  
DB 10792 AGTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10851  
QY 3749 GTGACCTCGGCGCTGAGCGCGCGCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3808  
DB 10852 GTGACCTCGGCGCTGAGCGCGCGCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10911  
QY 3809 TCAGAACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3868  
DB 10912 TCAGAACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10971  
QY 3869 CTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3928  
DB 10972 CTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11026  
QY 3929 GTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3988  
DB 11027 GTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11086  
QY 3989 CTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4048  
DB 11087 CTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11146  
QY 4049 CTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4108  
DB 11147 CTTTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11206  
QY 4109 GAAATGAGGCTGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4168  
DB 11207 GAAATGAGGCTGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11266  
QY 4169 GATTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4228  
DB 11267 GATTCGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11326  
QY 4229 GGGCGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4288  
DB 11327 GGGCGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11386  
QY 4289 GTGATGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4348  
DB 11387 GTGATGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 11446



Accession	Source	Organism	Gene	Map	Size	GC	GC%	GC3%	GC4%	GC5%	GC6%	GC7%	GC8%	GC9%	GC10%	GC11%	GC12%	GC13%	GC14%	GC15%	GC16%	GC17%	GC18%	GC19%	GC20%	GC21%	GC22%	GC23%	GC24%	GC25%	GC26%	GC27%	GC28%	GC29%	GC30%	GC31%	GC32%	GC33%	GC34%	GC35%	GC36%	GC37%	GC38%	GC39%	GC40%	GC41%	GC42%	GC43%	GC44%	GC45%	GC46%	GC47%	GC48%	GC49%	GC50%	GC51%	GC52%	GC53%	GC54%	GC55%	GC56%	GC57%	GC58%	GC59%	GC60%	GC61%	GC62%	GC63%	GC64%	GC65%	GC66%	GC67%	GC68%	GC69%	GC70%	GC71%	GC72%	GC73%	GC74%	GC75%	GC76%	GC77%	GC78%	GC79%	GC80%	GC81%	GC82%	GC83%	GC84%	GC85%	GC86%	GC87%	GC88%	GC89%	GC90%	GC91%	GC92%	GC93%	GC94%	GC95%	GC96%	GC97%	GC98%	GC99%	GC100%
4349	GGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	4408																																																																																																							
Db	11447	GGCGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	11506																																																																																																						
4409	GCGGATCCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	4468																																																																																																							
Db	11507	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	11566																																																																																																						
4469	CACCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	4528																																																																																																							
Db	11567	GACGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	11626																																																																																																						
4529	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	4588																																																																																																							
Db	11627	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	11686																																																																																																						
4589	GAGGAGAGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	4648																																																																																																							
Db	11687	GAGGAGAGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	11746																																																																																																						
4649	TCGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	4708																																																																																																							
Db	11747	GCGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	11806																																																																																																						
4709	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	4768																																																																																																							
Db	11807	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	11866																																																																																																						
4769	GAGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	4828																																																																																																							
Db	11867	GAGCGCGCGCGCATCTCGTCACTTTGCGCTTGTGTCGGATCTCGGTGACCACTTGTCCGG	11926																																																																																																						
4829	GATGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	4888																																																																																																							
Db	11927	GATGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	11986																																																																																																						
4889	GCACTGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	4948																																																																																																							
Db	11987	GCACTGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	12046																																																																																																						
4949	GCGGATCCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	5008																																																																																																							
Db	12047	GCGGATCCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	12106																																																																																																						
5009	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	5068																																																																																																							
Db	12107	GTCGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	12166																																																																																																						
5069	GATGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	5128																																																																																																							
Db	12167	GATGAGCGCGAGTGTGCACTGCGCGCGCATCTCGGTGACCACTTGTCCGG	12226																																																																																																						

AUTHORS Anzai, Y., Saito, N. and Kato, F.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-AUG-2002) Yojiro Anzai, Toho University, Faculty of Pharmaceutical Sciences, 2-2-1 Miyama, Funabashi, Chiba, 274-8510, Japan (E-mail: yanzai@phar.toho-u.ac.jp, Tel: 81-47-472-2072, Fax: 81-47-472-2086)  
 FEATURES Location/Qualifiers  
 SOURCE 1. 58343  
 /organism="Mycobacterium tuberculosis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:28040"  
 complement(<1. .910)  
 /gene="mycP"  
 complement(<1. .910)  
 /gene="mycP"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57022.1"  
 /db\_xref="GI:28394149"  
 /translation="MSPSGVEVLYDLIKRTVSNPIYODATHVAGLITTEAFVEARESGSDYPTVAHTMIQKRLNNLOCHYESLRGVGVDTLGVWKGACITARGILKAYDVADRTPWVADSFOGEPKATTDQHPMDKEMMLQYNEAVDLPISLATVQNFSEYIGLLDQDRVLPGMFDQTPMTAPFERTALVRNDGDSYGATMDVLTAAVPLRSPGFATIDYCIACGAEVAHEYRDRHGISDEIYDRGVYWRSLGILTPVHCRRSRSGSPPTGACHHITGSSGPRSNIPSLHRSGVSIPIRSRI"  
 1490. .2641  
 /gene="mycCI"  
 1490. .2641  
 /gene="mycCI"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cytochrome P450"  
 /protein\_id="BAC57023.1"  
 /db\_xref="GI:28394150"  
 /translation="MDRTCAMALPEQVAFSPORATGWPFAKWDGSPMTLVSRYEYHVALYVDPPVDTROPRLSPRSDGDSGFSMLLDPEHTRLRPMFISAFSVQVETMRPEIKETVDGIDRLRLALRPNDILTHLALPMSVQICHLGVPYRDRSPQERSRELASRPBKKDSMPALISLVEYLDGLVRTTANPDTGGLGATATERLLKGEITTHQELVNNATVLLAAGHSTASQNTGLSVLTLNHPETAELRQPELPMNAVDRLRTHSLADGILRRATLAIIVLGDHTIRAGDGLITLSSANHDINTGAELTFDIRPARHVAFGVPHQCIGLQNLARLEMEVTLTKLFRRVPLRLAQEPDALARVQSGSPIFGIDELVEM"  
 2662. .2886  
 /gene="mycCII"  
 2662. .2886  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ferritin"  
 /protein\_id="BAC57024.1"  
 /db\_xref="GI:28394151"  
 /translation="MRIVDARCVGAGCGCENTAPELFTQDDGLGLVDRPVTPEILLGPAREVDRCPVVALIRIESSVRTGMARGAG"  
 2883. .3866  
 /gene="mydI"  
 2883. .3866  
 /gene="mydI"  
 /codon\_start=1  
 /transl\_table=11  
 /product="4-keitolactonase"  
 /protein\_id="BAC57025.1"  
 /db\_xref="GI:28394152"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 3887. .5086  
 /gene="mycS"  
 3887. .5086  
 /gene="mycS"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57026.1"  
 /db\_xref="GI:28394153"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 5086. .58343  
 /gene="mycCII"  
 5086. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57027.1"  
 /db\_xref="GI:28394154"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57028.1"  
 /db\_xref="GI:28394155"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57029.1"  
 /db\_xref="GI:28394156"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57030.1"  
 /db\_xref="GI:28394157"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57031.1"  
 /db\_xref="GI:28394158"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57032.1"  
 /db\_xref="GI:28394159"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57033.1"  
 /db\_xref="GI:28394160"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"  
 58343. .58343  
 /gene="mycCII"  
 58343. .58343  
 /gene="mycCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="methyltransferase"  
 /protein\_id="BAC57034.1"  
 /db\_xref="GI:28394161"  
 /translation="MTRDRMANRSVVVTGALGFIGSHVEELIARADVTLGYSHIRAYDELSADHRVLLPVDDLDERRLQVFEHRSAGQITVICALIDGNAAYKRAHSAELITLANSRVSNLNCVDRGVEDVTLSSAAYCGPTSPAREDDARCTVSGENGQIVSLKLTIELIAEOYRQGVGFIIVRPAVNVGVRPGSPASRVIPAIATAESGSEETIKMDGQQTSSPYVTVDIVRALALVEYKPHSLNVTDDRYTSMULIARVAFSTGRTARLHHKPAQVPGVAGVLDITTRRREVVDTTPRLTEBGLBETVRRTRHIGR"

```
/transl table=11
/product="methytransferase"
/protein_id="BAC57026.1"
/db_xref="GI:28394153"
/translacion="MTAQTEFEDATVQDVRLAGHDSLELRELTOKXDPMLSRLLVA
EILSRCPSPNDTPVVELAVHGSERFRHFLRVNDSPIRPGADEGFGMLVEVEL
TELLRELFVETBRPAGVGTGKLPYLTDBEAEVQIGTYLTAQGGYKHKEMGGSL
KPDLSLSRFTPKFGFLHMTPLPHDRHRRDYNQOVRLEYLIGVGGYKHKEMGGSL
BMKSFPPGQIYGLDLMOKSHVDELRIITQGDNDADRLDIARXGFPDVIDDG
SHINAVRISFALPHVHFGGILYEDMTATWPGGQADQOESGSLGLKSLI
DAIQOELPSDPNRSFGYDNRNIVGLHVNNAFVEKGRNDEG31PTWIPRDESLVQ
ASSGGAT"
5181..6443
/gene="mycd"
5181..6443
/gene="mycd"
/transl_start=1
/codon_start=1
/product="glycosyltransferase"
/protein_id="BAC57027.1"
/db_xref="GI:28394154"
/translacion="MRITALTIGSRGDVOPFLAGVGLARGHDTILGADEGLKHLVE
RAGLTPTPGDGFPTMEVVDLSRSPYLDLKLPAVEEYDOQVVDIILAG
OGADAVVYPLITAAVGSADVMAAANMPNTRTPRPAVEGQRLGFLPLHLF
THARAAREAMRPERVDYRROGLEPFGSGPRLIGRERYLYPSPSVLPKRD
WPARAVTGVFWMDOMOPTRLEADGAPPAVAVHGGFGLTAHREOYLTAIAA
RAAGRIIVVDGDPVSDVYLHVDYTRMLEPRTAAVHGGFGLTAHREOYLTAIAA
VVVFPADHPFMARLLEIGVAPADVFLTRALVASTVAVKVRBDSIGARARAPH
AVRERGVENACCVSWIDNSRTLTG"
6720..19643
/gene="mycal"
6720..19643
/gene="mycal"
/note="polyketide synthase"
/codon_start=1
/transl table=11
/product="protomycinolide IV synthase 1"
/protein_id="BAC57028.1"
/db_xref="GI:28394155"
/translacion="MVVPSVTAPESSVALVGSCLPGASVGEFMDLTTEGDAVAR
OPDGSWRALAEFGTDFAGFGMSPROAAADPOQLVLELGEALHEHGVPAITAG
SRGVGVVADYATSLWRGTTPDGYTPGLNRMILNLSYLELKGSPVVDAS
OSSSLVAHLACSLLRGSDAIYAGVSLISTPESTAAASGLSPGRCPTPAR
ANGVYRGGGAAYVLRSLSDAIVDGRVAVHILIGAVNNDGSGELAPRAQEEYV
RRAVDRSGVPAQVITVELHGTTPVGDPEALCALAVNGADRADPLHVSXKNI
GHLGGAIGTGLKAVLCVREGVLPPLNHPNPGIPLAQLRLVQTVVORMPDAD
BGRRIIAVSSFGKGTNAHLVLAQAEBAVRPAVSPTEELPAVPLVLSAREALAGQ
AERLEHLDHRLPLDVGFSLATRDVPHRSVRADGEPVAAGSGLAEGRVAVY
PFGQAGQMTGMAAASLTESPVPAARLTGCAAVTPPVMSILDLVLAHVGSLPDRDD
VVOACPAVMSLAELMRSGVPRVAVHGSQETIAAVYAGKISLEDARIYALRSQ
LITRLHIGAVSVGLPVAETERLITNMSGRLHLAVNGPNVAVAGDGLDELVT
ECERRQARVRAVEVASHPHYERIRDELMTALRVTPQPSVAVSTADVCIITGBE
RLDAGTWNHLKQVRFADTLQMSDTGTSFVLSLHMLVLALESLIAOETTALCQ
VYGLSLRDEGGLRMLASAEAGHGLPDMRKVFPVGTGHEVELTEYLAFORVYAP
DAERTVPPAAVDAVYRPAASPEQATPAQATTCQYVDLITRLVLAIVQGLPSMET
VPODLITPKQGLSEYAAVERNRRLARVGLSLPATIYTPPTTAVAGELGDLHRE
SOLVITVPLVDEPILAVMGCRPFQGVSPEDLNDVVASVDVSPVPRGMPLRE
SLFDGDPAGITTYCREGGFLDAAFGPFPVSRREXVANDPOQLLEHGVREVG
AGFSAGVGRGRTVYAGVMOQYGPRLYESAEHRGGLITGTAIVSGRVYLG
LEGPATVDTACSSLVALLHVAQALRSECDVLAGVMAVTPGIPIEFSQGLTA
VDRCKAVSDSADGTGMSGVGVVYVQVRLSDAVALGRVLAVRGSAVNDGASNLT
APSGSQQAVITRAALASARLSDVDVVEGHGTIRLDITLQALATYGRRGGSLL
LIGSVKSNIGHQAAAGVAGVAMVLAMRGGVPAALHACAFACQDMSGAVSAYVGG
EACAMPESTRPRAAASPGVSGTNAHVLBOAPVVPDASSGAPVAVASABSSS
VSPVWLSAREWGLASAGRLAELASGLSGSGSGGVAGVGDVMSLIGSRSVFMR
GVSLSGVRLGYLAGGSGSVVVGSGVSGSVVPGQAGQVMSLAKGLIGESG
FSBAEDVGVGVGLVDMVPRGACREVRGDLGRDVQITSPVMAVLAALRSBSG
VYPRGVHSGQETIAAVYSGALSVBGRVAVYGRVIGARLGGAGANVVGASGV
VGLLGLGGLGGLAVAVNGGVVAGDGKGGVGAAGVAGVAGARLWVVGASGV
GLVDVDRDLGLVLEGRGSSVTPYSSVDVGVGVGGLDGGVPMNRLRLRPFEB
VRLVGDGSSFEVESAHPVLVGAIEDVGEVGVGFGVAGVAGVAGVAGVAGV
AAGVVRGLPVQMRVPLVAGRGVGLPTYPFOHDDYMLAPARTVAAAGATYAADGA
AGVGMRYGVKALPVPTPRLHGRVLAHGHATGVDLDSGVVADILTRHAGTTEHL
```

Query Match	14.6%; Score 1309; DB 1; Length 56343;
Best Local Similarity	63.6%; Pident. No. 3,7e-85;
Matches 1952; Conservative	0; Mismatches 1107; Indels 8; Gaps 2;
QY 2047	GGGCAAGGGGTCACCTTCTGTCGGTGCCTGGCGCGCTCGAGAAACCAAGTCCGC 2106
DB 48080	GTGCGGTGTCGGCTCAGACGGGCAACCGGGGTTCGACCGGTGGCGGTACCGGACGAGAA 48139
QY 2107	CTGAGAGATGATCGCCCGGTACGGTCACTGCGCGGTGACCGGACGACGACGACGCTG 2166
DB 48140	GGGAGAGAGGTGTGAGACGGAACACACTCTGTACCCCGGCGGATCCAGAGACCGG 48199
QY 2167	CCTCACTTGCACGATGCTCTCGGACCCGGGCACTTCGGGTCTGGCGCGGTCAACC 2226
DB 48200	CAGGGCAGTTCAGAGGGCTCACAGCCTGTTGAC--CGGACGAGATATCGGATCCG 48257
QY 2227	CCGGCGGATCCGGCTGAAACCCCTTGGCTCCGACGGCGGAGNNNNNNNNNNNNNNNN 2286
DB 48258	CAGTAGCGCGGCAACTCCGCTCGAAGGCTCGATCTCCGGGCGGACGAGTACACGCG 48317
QY 2287	NN 2246
DB 48318	GAGGCGAGACCCCGGCGACAGGCGCGGTGATCTGTCGGCGACCTCTGRTAGGCGGCT 48377
QY 2347	NCNTCNCCNCCCGCGCGCGGGGGGTTCAGACCGGCTTCAGACGCGCGGATTCAGGCG 2406
DB 48378	CCGACGTTCAGAAAGGCGACGTCGGGGGTGTGGCGGACACACGCGTCAAGCGAGGAAGCC 48437
QY 2407	CCGATGCTATCCCAACCCGTCCGACGTCCTCGCTTGAATTTGTTAGCGCGGCGGTCAAC 2466
DB 48438	ACGAGGCTACCCCAACCCGTCCGCGGTGTGGTGAAGCTGTGTTCAACCGGCGGCTGTAC 48497
QY 2467	CGACTGATCGAAGACGTCGAGGAAAGAACTGTCGCCCGGCTCGGATCGATGTCGCGGCG 2526
DB 48498	GGCTGTGTGAAACCGTTCATAGATCTCTCCCGGCGTGGGCGGCGATCCGCGGCGCC 48557

QY 2527 CGAGTGAAGAGTCTGTGGAAGAACGAGTGAAGGCTCCGGTCCGGGGTGAACCCGACCGC 2586  
DB 48558 GCTGGACGAGAGTTCCGAGCACCTCGGCGAGGAGGTGTCCGAGCCGATCCGAGCGGC 48617  
QY 2587 GACGTAGCGGGTGGCCCCCGCAGCCCGGGGAAAACCGGCTCCCGGTACAGATTAGAGTGC 2646  
DB 48618 GATGTATCCGGTCCGACACCGGCGAGGTCCGGGAAGCCGGCTCAACGTATGAGATGATGTC 48677  
QY 2647 TCCGAGCAGGTCCATCTGACACCGCACCTTGCGATTGGGCGGTCCGATCGTTTCCGG 2706  
DB 48678 GCGCGCAGGTCACCTGACACCGGATCTGGGATGGCGAGCCGGGCGATGTGTCTCCGG 48737  
QY 2707 CCGGATCCGGAACAGCTGGGGGTGACACCCCGAACCGAGGCTCTCAACCGGTAAACCG 2766  
DB 48738 CCGGATCCGGAATATCTCCCGGTATCATGCCCATGGATCTCAAGCGGTAAACCGAA 48797  
QY 2767 GTCGATCTGATGCGCCCGGGGTCCGCTCGGCGCGGTAGTCAAGAACCGGGCGAGGCCCTC 2826  
DB 48798 GTCCATCTCAAGCGTGGGCGCCCGGTGAGGGCCCGGCGCTCGAATCTCCGAGCGAGTC 48857  
QY 2827 CCGGCACTCGGTCCGTTGCGCTTCGACAACTGCGGTCTGTCGACCCGTGTATGTCCTC 2886  
DB 48858 GCGCAGTTCCGCGGTTCTGTGGGGGAGAGCGCGCGTCCGATCGGCTGTGTGTCTCTC 48917  
QY 2887 GCGGACGGGTACGAAGTCCAGCGGCGCGGTGGGGCTGGAATGTTGAGTTGGGATGAA 2946  
DB 48918 GCGTACGTGAGAGAGTCCGATCCGATCCGATCGGCTGTGTGAGAGTCCGAGTAA 48977  
QY 2947 GTCCAGCAGGTGATGATGAGCGGTGGGCGCGCCCGGAGAGATGATGTGTGAGAGCCGAG 3006  
DB 48978 GTCCAGCAGGTGATGAGAGCGGCGCGCGCGCGCGAGCAGAGTAAAGTCACTCCGAG 49037  
QY 3007 CCGGACCGGTGCGCGCGGTTCGGCGCGCATCCGAGAGATCCCGAGAGTTCTTCTGAC 3066  
DB 49038 CCGGACCGGGCTGGCGGTTCGGCGCGCATCTGAGAGTCCGAGGTTGGCGCGGAC 49097  
QY 3067 CCGTTCCAGAGGCGCGCAGCTTGGCGGGTGGGCGGTGATCTGTGTGTTCAAGCCGTA 3126  
DB 49098 CCGCTCCGAAACGCTTGGCGGTTCGCGGTGGCTCTGATCTCTGTGTGTGAGAGCGTA 49157  
QY 3127 CAGGAGGTACGAGACCGGCGCGAGGTCCCAAGGCGGGCTGGCGCGCAGCGTCTGTC 3186  
DB 49158 GAGGAGCACCCGGAACGGGTGACAGTCCCAAGGCGGGTGGTGGTCCGACGCTGTGTC 49217  
QY 3187 GGTAGGCGGAGGCGGTGGTGGTGAAGGTCAAGGCGAAGCCCGCGGGCGGGGTGCGA 3246  
DB 49218 GGTAGGCGGAGGCGGTGGTGGTGAAGGTCAAGGCGTGTCCCGTGGCGGGCGCGCT 49277  
QY 3247 CACCAACGCCCGGATACCCGGGGTGGTGAAGTTCCAGGGCGCGCGAGAGGTATCATGCG 3306  
DB 49278 CCGCAGCGCTCCGACCTTCCGGGTGGTCAAGCGCTCCAGCCCGCGAGAGGTATGTTGC 49337  
QY 3307 CGTCGGGTCTCCGACCGCACCTGTCGATGACCGAGGTGAGATCTTGTTCGGGCGAC 3366  
DB 49338 GTCCGAGATTGTCCCGGGGCAAGTGTGATGACCGAGCGAGGGGGGTGCGCTGTTC 49397  
QY 3367 CACGCACTCCGCTGTACTTGGACCTGTCAACCGTACCGAAGATGACAGCGGAATAT 3426  
DB 49398 CAGCAGGAGCGCTGTAGCGCGCCCGGTCAACCGGACACAGAGTGGCAGCGGAATAT 49457  
QY 3427 GCACTCGGGCGGGGTGAGACCGCGACCTGTGATCCGGAAAGGGGGCTCCCGCGTATCGC 3486  
DB 49458 GCAAGTCAACCCGGGTGAGAGCGGACGATCCGGGGGAGAGACCGGCTCATGTCCGAGCGC 49517  
QY 3487 AGCTTCAACCGCGCGCGCGGTTCAGCGGCTGTGATCGTGTGCTCAAGTATTCGCGCGC 3546  
DB 49518 GCGCTTCAACCGCGCACCTGCGTTCAAGCGGAGAGACCGTGTGGCCGCAATTCCTGCGCGC 49577  
QY 3547 CCGGCGCTTCAACCGCGCGGTGCGGCTCCAGCTGGGAGTGGATCAACAGGTCAAGACGCTC 3606  
DB 49578 CCGGCGCGCTTCAACCGCGCGGTGCGGAGCTCGGAGCACTGGGAAAGATGTCCAGCAGCGC 49637

QY 3607 GCCGAACCGGCGCCGGTCCAGGTCCAGACCGGCGCATCTGCTCAACGGGGTGAACGG 3666  
DB 49638 GCGGAAACGGGGCGCGGTCCAGCGGATGTTGTGTGGCGGCGCTGTCTCCAGCAGGGTGAACGG 49697  
QY 3667 CGAGCGCGGTATGGGCGGTGAGTTCCGCGAGCCCGAACCGCTCTGTCTTCCGCGGTTC 3726  
DB 49698 TCCGTTGCGCAAGATCAACCGGCGAGCGGAGAGACCGGACCGTCCGCGGTTCGGG 49757  
QY 3727 GGAATGGGCAACCGGTGAGCGGGTGAACCTCGGCGCTGAGCGCCGCAACCGGCGCGCG 3786  
DB 49758 TTGCGTCAAGCCCAAGACCGCGGTCAACCGGTACCGAGATCTGTGCTCCGCGGTGAG 49817  
QY 3787 GTCGCGTCCGGTCTGTGCGCGGTCCGATCTCGTCCGATCAAGCGGTCGCGGTGAGAGT 3846  
DB 49818 GTCAAGGTGGGTCTGGGTGAGGTCTGATGACGAGCTGATGTGGAGGTCCGAGGATCC 49877  
QY 3847 GGTCTTCAATG-----ACGTGGAACCTTCTGGGCTCTGTGTGAGAGATCAAGAAC 3900  
DB 49878 CATTCGCTTGTCCGTTCTGTGGAAGTGTCCGCTGGCTGTGACTCGGAGGTGGCGGCGC 49937  
QY 3901 CGTTGCGTTTCCGCTGTCCCATCCGCGGTGATCAACGACCGCTGTGTGGTGGAGTGC 3960  
DB 49938 GCGTGAAGCGCACTGTGACATGACGGGTGATCTCCGACCTGTGGGACACGCGAGCC 49997  
QY 3961 GATGAGTTCGACACCCGCGGAGTGTCTCACTGACACCGAGGGCGGTCGAGAGG 4020  
DB 49998 GCAAGATTTGACACCCGCGGAGCGTCTCTGTGGAGACCGTTCGCGCGCGGTGGCAGCG 50057  
QY 4021 CAGACACCCGTTGGCGAGCCATCTCGTCTGTGTCAAGCCGACGGGTCTCGTGGGT 4080  
DB 50058 CCAATGACGTCGCGCAGCGCGCTCGGTGGCGAGAGACACCCGTGGCGCCGAGCGGT 50117  
QY 4081 AGGCGCTCATCTGGTGCAGCGGGGAGAGTAAAGGCTGTGCGACGACTTCTCCGCGC 4140  
DB 50118 AGGCTGTCTGTGTGACACCGAGGGAGAGTAAAGGCTGTGACGACTTCTCCGCTGC 50177  
QY 4141 GCAGATCGCTCGAGCGGTCACGATGATGCGGTGGCGGTGCTCCACAGATGA 4200  
DB 50178 GAGACAGGTGATCAGCAGGTCGCGGTGAGCGCGCGGACCGCGGGTCCATCTCCACA 50237  
QY 4201 TCACTACTGTGATGTCTCTCTGTGTGGCGGAGCGAGTGAACCGCGGTGA 4260  
DB 50238 CCACTGATGTGATGTGCGGTCTCTCGGCGTGAAGGTGATGTGTGACGCGCGGAGCA 50297  
QY 4261 CGTGGCGACGCTGCTGTGTGAGAGCGGTGTGAACCGGATGTGCTCTGCTCGG 4320  
DB 50298 GCGGGCGAGTTCCGCGCGGTACAGCTGTGATGCTCGCTGTGTGTGCGACAGACTGTCT 50357  
QY 4321 CGAACGCGTCCAGGAGTGAAGCCCATGAGCCGCGGCGCACTCGCTCATCTTGCCTTGG 4380  
DB 50358 CGAAGGCTCCAGGAGGTCAGTCCATGGCGGCGAGGCTCGCTCATCTTGGCGTTGG 50417  
QY 4381 TGCAGATCTGATGACCACTTGTTCGGGCGCGATGCCGAAGTTGTGATGGCCCGGATCC 4440  
DB 50418 TCCGCGCCCGCGGTTGTTCGCGACGAGCCGGAACCGGAAGTTGTGAGGGCGGAGACC 50477  
QY 4441 GTTGGCGCAGAGCGCGGTGTCGATGACACCGCGCGCGCTTCGAAGGCGGTGACCGCT 4500  
DB 50478 GGGCGGAGAGCGCTGTGTGTGTGAGATGCGGCGCGCTTCGAAGAGTTTCAAGACT 50537  
QY 4501 TGTGTGGGTGAGACTGAACACTTCGCGGTCAACCGGATCCGGAACGTCACCGC 4560  
DB 50538 TGTGTGGGTGAGACTGAAGACTTCGCGGTGCGGAACCGCGCTGCGCGCGCTTGG 50597  
QY 4561 TGTGTGAGCCCAAGGCGGTGGCGCGGTGAGAGAGAGCTTGAACCTGTGTGTGGCGGGA 4620  
DB 50598 CCGTGCACCCGAAAGCGGTGGCGCGGTGAGAGAGAGAGGTGTGTGTGTGTGG 50657  
QY 4621 TCTTTCGACGAGCGCTCAACAGTGTCTGTGGCGCCCAAGGTATACGCGGATTCGCGC 4680  
DB 50658 TGCCTGTAGGCTCTTCCACCGCGGAGGACCGCCGAACAGGTGACCGCGGATGAGCGG 50717  
QY 4681 CGGTCTGGGGGTGAACGAGCCGCGGACGTTGTTCGGGTTCGACCAAGACCGGTTCGCGGCT 4740

Db	50718	AGGGTCGGGGGGGTGATGCGCGGCCCGGACCCGCTTCGGGGGTGGAACAAACCGGTGCGCGGGT	50777
QY	4741	CGACGTGCGCAGAACACCGGTGTGATGTCGAGCCAGCTTCGCGCGGTGCGCGGTGCGCGCA	4800
Db	50778	CGAGGTGCGCAGAACACCGGCGCGCAGGCCACGACCTGCGCGGGGTGCGCGGTGCGCGCA	50837
QY	4801	AGGTCATTGACCGGCATATCATCTTACACCGGTGAAGTCACCGGCCCGCAGACCAAGTTTCCA	4860
Db	50838	AGGTCATTGAGAGGGTACATCATCCTCGCGCGGTGATTTGATGATGCGCCCGCATCAGCATTTGCA	50897
QY	4861	GGGAGCAGGTGAGCGTTGACAGGTGCGCATGTGAGTGCCTGACCCCGACCAAGTTCGCGCAACC	4920
Db	50898	GGGAGCACCGTGGCGCTTCGAGGTGCGCAGACATGTGCGCGAACCCGACCCCGGTGCGCGCATCC	50957
QY	4921	GGGCGTTCGAACTCCCGCACAAGAGGGGCCCGCGTTGGTGAAGCAGATTGTTGTAAGGAGCC	4980
Db	50958	GGTCTCTGAACTCCCGGGGTGACGCGGCCCTCGGTGGTGAAGCCACTGGTGTGTCCAAAGCCC	51017
QY	4981	ACTTCAGAGCGGGCAGGAACCGTCTCGCGGTGCGCGATCGTGGCCGCGGCCCACTGTGACAGG	5040
Db	51018	ATGACACACCGGTGGAAGAACCGGGGCCCGGTGCAAGGGGCGCGGCCCGCCGAGCGTGAAGGG	51077
QY	5041	GGTCCAGGAACGCTTCGGGACCGGCCGAAGTCCGACAGATGGTTCGGTAAGCGCTTCAAGC	5100
Db	51078	TCTGGAGAAAAGGCTTCCCGGCCCGCGAAGAGGGCCAGGTCTGTGATGTTGCGCTTCAAGG	51137
QY	5101	CGTGGCC	5107
Db	51138	CGCTGCC	51144

3-reductase"  
/protein\_id="AAFS9931.1"  
/db\_xref="GI:7329191"  
/translation="MPSRLRFVGLGADILRLRTPTALHEDVTVAVSSRDTARRA  
AFNAFCEAVPGHQALDDIDALVPLPVNWHPWEEAALLRGNHVEKPLDRA  
TRSGEDILALARSGLVIMENFTSLHIAQGTVDLIDDTIGELSLSAFTLPP  
EGDRLVYDPPGGLALDIGIYPLRALHPGPIPLAAGAVLRERRNVYSHVLL  
TPHVAIVELAFGEMHAYRSRYTLRGTAGRLRLORAPLPTPTAHPRIHQDADIV  
DLPPDQANLVKDPVLAIVRSGRLTQHDADVAHQDLVERVAVARVRC"  
complement(1173..2597)  
/note="Oxid; involved in the 2-deoxygenation step in  
dTPD-L-oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=11  
/product="dTPD-4-keto-6-deoxyglucose 2,3-dehydratase"  
/protein\_id="AAFS9932.1"  
/db\_xref="GI:7329192"  
/translation="MIMGILPAMSEAMGSVPTAGSEVSTCALLSMILDARRRALTG  
HYPFELSGMOEPDENTGILRLTSGRSPSIEGLRYRTHCMGMSWTPILVOPETLIG  
LLVRFDILLVQAKNBERNIGGLQLSPVQATRNTRYRNGGCVRLIEFASRR  
GGRVTLADVSEGGSPFLKRRNRMEVEALDDVPLDDDPFWLSLGGRLTLKPLH  
NDKTVLSTSCLPDPADPGRQSPAAPFAAIVYSLTRGATVLAHTMGRLIGMTLDR  
SRRLVQOORVLEBTAFSGMRDRDHAIHAKQGDYFVIGSVASRSVSSWSPPLA  
PVGPEGLFVRRILRGVTLVHLHARTAGLXGNGEEMAPVQCRLNRAVPAEKYRA  
LDVYLSADPGIRYDTLOSBEGRFHHAENRYVVEABDPPEVPPDPMTLTHQL  
ALIHNSVYVNSLVACIALS"  
2933..3541

RESULT 14	AF237894S1/c	9523 bp	DNA	linear	BCT 26-MAR-2000
LOCUS	AF237894S1				
DEFINITION	Streptomyces antibioticus T99 oleandomycin biosynthetic gene cluster, partial sequence.				
ACCESSION	AF237894				
VERSION	AF237894.1	GI:7329189			

```

/product="qtDp-4-keco-6-deoxyglucose 3,5-epimerase"
/protein_id="A0F5993.1"
/db_xref="01:7329193"
/translation="MELLDVNGALVYTPETMRDERGSEFLFMRGRTQKRTIGPLSLA
QANCSVSKRGITGRIFPADAPQGAQRKASGVNLDVVVDYGRSGTFCRMAVRID
AARQGLVLAGLGHAFMALIDDAITVYLISQPVAAERARVLDPLDAIGIEMPTDID
IVLSKDTNATSLAQAATGILPDIPEBCRAYIAAAGVGPAA"
3538..4608

```

ORGANISM	Streptomyces antibioticus Streptomyces antibioticus Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE	1 (bases 1 to 9523)
AUTHORS	Disieger, G., Park, S.-H.H. and Floss, H.G.
TITLE	Mechanism of the 2-deoxygenation step in the biosynthesis of the deoxyhexose moieties of the antibiotics granaticin and oleandomycin
JOURNAL	J. Am. Chem. Soc. 121, 2611-2612 (1999)
REFERENCE	2 (bases 1 to 9523)
AUTHORS	Park, S.-H.H., Sohn, J.-K.K., August, P.R., Niggemann, J. and Floss, H.G.
TITLE	A cluster of genes from Streptomyces antibioticus involved in the biosynthesis of the deoxysugar moieties of oleandomycin
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 9523)
AUTHORS	Park, S.-H.H., Sohn, J.-K.K., August, P.R. and Floss, H.G.
TITLE	Direct Submission
JOURNAL	Submitted (22-FEB-2000) Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA
FEATURES	
SOURCE	Location/Qualifiers 1..9523

```

/codon_start=1
/translate_table=1
/product="GTP-D-glucose synthase"
/protein_id="AAFS9934.1"
/db_xref="GI:7329194"
/translation="MKALVLAGSGSTRLPDITTHSAKOLVAVANKPVLPGYGLAAIAA
GITVGLIVGTASRYAAVAGDGAKKRFLDITTYEQSPPLGLAAAVLTAHTYLDDEDDP
MYLGNFIVGSGIDDLVTRFDRGRPPAIRILLTVSDSAGVAELDDGKGVGLIEKP
RHPSDIALWGVYFPTPAIHAVALEPNSRSGEIVTIAIOHLINDGADIOSVIEKY
WKQDGNVAMLEVRVTALEDPLEPIEGTGVHTVIGRAVVGAGAVRTNRIKMPALI
GAGLISISVYTGPTSVGDNCRILGSEMERIMLASALIGVARIEGSLGRNVQYQ
SLHAPNHRPVLGHSKVEYQS"
4605. .5612

```

```

CDS      complement(167..1153)
         /note="Orf11, involved in the 2-deoxygenation step in
         dTDP-L-oleandrose biosynthesis"
         /codon_start=1
         /transl_table=11
         /product="dTDP-3,4-diketo-2,6-dideoxyglucose

```

```

CDS      5609..6493
         /note="orf, involved in the 4-ketoreduction step in
         dmpd-1-oleandrose biosynthesis"
         /codon_start=1
         /transl_table=1
         /product="4-ketoreductase"
         /protein_id="AAFS936.1"

```

/db\_xref="GI:7329196"  
/translation="MRWLTGSAAGMLGREIVRLAENEDVAAAGHDHLYTEPFV  
AALAEHRPGIVNCAATYAVDADETDAALALNABPRLAECAPHAGRLVHSTD  
YVFPGDARTPEADHPTAPRAYGRTDDEQATLTALPTATVLRPMILGRCRSY  
RMTERAERGAIDVADQCSQPTWGDADRIIAVGRHPGHEILHATNAGSTYD  
LAQVFRILADPGRVPTTGAARPRAPRAYSVLGHDRWGTGLAPLDMRSALRE  
AFPDILAEHPTRGAA"  
6490..7671  
/note="orf5; possibly involved in the 4-deoxygenation step  
in dndp-D-desosamine biosynthesis"  
/codon\_start=1  
/evidence="not\_experimental"  
/transl\_table=1  
/product="3,4-dehydratase-like protein"  
/protein\_id="AAF5937.1"  
/db\_xref="GI:7329197"

CDS  
7792..9249  
/note="orf4; possibly involved in the 4-deoxygenation step  
in dndp-D-desosamine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAF5938.1"  
/db\_xref="GI:7329198"  
/translation="MALQVDRTPLDPLIVCAVMPGNSAAAALPLIEBARGTLEP  
DLAEHLALAEVGTPEPLSEBRLGIDRAPTAVALAVPRTPLAEVAPRA  
GMTNTTILPERGVLDAAVREBPAPYSGLYPGSCMRCHFCVATGARVQQA  
LTDNAMPASLIDRMPDNPALVLSGLLEPLNPGDVAATTKGAFORVADNLTRE  
ALTGTDLRQPLMDLVALTSLVGLSDVDVATTKGAFORVADNLTRE  
EAPVLEGLNLTILPGRALITGLADYFADLMDAADPVDPLTLEEDSGPDKLAP  
BERVLEHGLAIFERIRTRAPSLHVQVGLVGLSDVDLPLRIEPTRPAPHP  
VAVQVDLADYVLRKAPFGLQCAERVVAARLTGTLEBSEVVRFTYEGNOVAPRC  
EYFLDGDQTVTALNOMETDIADGMAEHRGFLK"  
7792..9249  
/note="orf4; possibly involved in the 4-deoxygenation step  
in dndp-D-desosamine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAF5938.1"  
/db\_xref="GI:7329198"

CDS  
7792..9249  
/note="orf4; possibly involved in the 4-deoxygenation step  
in dndp-D-desosamine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AAF5938.1"  
/db\_xref="GI:7329198"  
/translation="MALQVDRTPLDPLIVCAVMPGNSAAAALPLIEBARGTLEP  
DLAEHLALAEVGTPEPLSEBRLGIDRAPTAVALAVPRTPLAEVAPRA  
GMTNTTILPERGVLDAAVREBPAPYSGLYPGSCMRCHFCVATGARVQQA  
LTDNAMPASLIDRMPDNPALVLSGLLEPLNPGDVAATTKGAFORVADNLTRE  
ALTGTDLRQPLMDLVALTSLVGLSDVDVATTKGAFORVADNLTRE  
EAPVLEGLNLTILPGRALITGLADYFADLMDAADPVDPLTLEEDSGPDKLAP  
BERVLEHGLAIFERIRTRAPSLHVQVGLVGLSDVDLPLRIEPTRPAPHP  
VAVQVDLADYVLRKAPFGLQCAERVVAARLTGTLEBSEVVRFTYEGNOVAPRC  
EYFLDGDQTVTALNOMETDIADGMAEHRGFLK"

ORIGIN  
Query Match 14.3% Score 1276; DB 1; Length 9523;  
Best local similarity 66.5%; Pred. No. 1,66-82;  
Matches 1937; Conservative 0; Mismatches 910; Indels 65; Gaps 5;

2395 GCGATCAGGCCCGATGCTATCCACCCGTCGCGACGTCGCCGTTGAGTTGTTGAG 2454  
Db GCGCAGGAACCGCGTCTCCGCCAGCCGTCGCGAGATGTCATCTCATCTGTTGAG 9187  
2455 CCGGCGCGTCAACGATGTTGAAACCGTCGAGGAAGAACTGTCCTCCCGGTCGCGATC 2514  
Db ACGTCCGTGACGCTGCTGTCGAACCGTCGAGGAAGAACTGTCCTCCCGGTCGCGGCG 9127  
2515 GATGCTGCGGCCGACGTGACGAAGTCGAGCAACGAGTCAGAGTCCTCGGTCGCGGAT 2574  
Db CACTTGCAGCCCTCGTGAACAAACCGCGGACGACCTCTCTCAATTCCTGTCGCGAT 9067  
2575 GACCCGCGCGGACGTAGCGGTCGCGCCCGCGACCGCCGCGGAAACCGGCTCTCCGAT 2634  
Db CAGCCGCGCGGACGTAGCGGTCGCGCCCGCGGAGCGCGGAAACCGGCTCTCCGAT 9007  
2635 CAGGTAGACGTTCTCGAGGAGTCGATCTGACACCGGCACTTCGAGTGGCGGTGCGGCG 2694  
Db GAGGTAGACGTTCTCGAGGAGTCGATCTGACACCGGCACTTCGAGTGGCGGTGCGGCG 8947  
2695 CATCGTTCCGCGCGGATCCGGAACAGCTGGGCGTCGACACCCCGACGAGGCTCTCCAA 2754  
Db CATGCTCTCGGCGCGGATCTCGAGGATCTCGGCGTCGAGCGCCGACGAGGCTCTGCA 8887  
2755 CGCGTAACCGAGTCGATGCGCCGCGGATCTCGGCGGCGGATGTCGACGACGCG 2814

8886 GCGGTAGCCGATGTCAGCGTGAAGGACGCGGCGCGGTGTCGCGATCCGTTCTCGAAAGC 8827  
2815 GCGAGGCGCCCTCCCGCAACTCGGTCGCTTCGACCTCGGAACAACCTGCGCTGTCGCCAGC 2874  
Db GCGAGGCGCCCTGCTCCAGTTTCCAGCGCTTCTCGGCGGCGAGCTTGCGCTGCGGCGCTCC 8767  
2875 GCTGTAGTCTCTCGCGACGCTGACGAAGTCGACGCGCGGTCGCGGCTGGAATCTGTTGAG 2934  
Db GCTGTAGTCTCTCGCGAGGCTGAGGAAGTCACCGGACGCGTCGCGGCGCGCTGTTGAG 8707  
2935 TTGCGGCAATGAATGCAACGATGCAAGGCGGTTGCGCGCCCGCGCGGAGATGATGTG 2994  
Db GTGCGCAATGAATGTCAGCGCGCGGTCACCGCGCGCGCGCGCGCGAGATGATGTA 8647  
2995 GTTGAAGCCGACGCGGACCGGTCGCGCGCTTCGCGCGGATCCGAGAAAGTCCCGAG 3054  
Db GTTCAAGCCGACGCGGACCGGCGCTTCGCGCGGATCCCGGCGGAGCGGATGAG 8587  
3055 GTTCTTCCTGACCCGTTGAAAGCGCGACGCTTCGCGGTCGCTGATCTGTCGTT 3114  
Db GTTGTCTTGAACGCGTGAAGGCGCTTCTTCGTGATGTCGTCGTCGTC 8527  
3115 GTTCAAGCCGTCAGGAGGTCGAGACCGGCGCGGATCCCAACAGGCGCGGTCGCGGCG 3174  
Db GCTGAAGCCGTCAGGAGGTCGAGAGGTCGAGAGGTCGAGAGGTCGAGAGGTCGAGAGG 8467  
3175 CAGGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3234  
Db GAGGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 8407  
3235 GCGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3294  
Db GCGCGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 8347  
3295 GAGGTATATCGCCCTCGGTCCTCCGACGCGACCTGTCGATGACCGAGGTGAGATCTC 3354  
Db GAGGTATATCGCCCTCGGTCCTCCGACGCGACCTGTCGATGACCGAGGTGAGATCTC 8287  
3355 GTTCCGCGGCGCAACGACCTCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3414  
Db GTTCCGCGGCGCAACGACCTCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 8227  
3415 ACAGCGGAACATGACGTCGCGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3474  
Db GACCGGAACATGACGTCGCGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 8167  
3475 CCGCGGTCACGCAACCTGTCGACGCGCGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3534  
Db GCCCTTACCGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 8107  
3535 GTACTGCGCGCGCGGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3594  
Db GTACTGCGCGCGCGGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 8047  
3595 GTGAGCAGCTGCGCGAACCGGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3654  
Db GCGCGAGACGCGGCGGAGAGTCGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 7987  
3655 CCGGTCGACCGCGAGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3702  
Db GCGGTCGACCGCGAGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 7927  
3703 GACCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3762  
Db GTGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 7867  
3763 GAGCGCCGCGACCAACGCGCGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3792  
Db GCGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 7807  
3793 GTGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3846  
Db GACTGAGTGCACGCGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 7747



QY 3847 -GGTCTTCAATCGAAGTGCAGAACCCCTTCTGCGCTCTGTGAGTCGAGAGATCAAGAACGTTG 3905  
Db 7746 GGGGTCTCCCTCCGGAGAGTGCAGCTCTCTGCTGCTCCGCCCTGCTCCGCGAGACGCGCCG 7687  
QY 3906 CGTTTCCGCTTGTCC-----ACTCCGCTGTGATCAACGCAACGCTGTGTGTGC 3955  
Db 7686 GGGCTTGCACCTGCGCTAGAGCGCGCGAGCTTGGGCGGTGATCCGCGGCCCTGCGCGCGGCG 7627  
QY 3956 GAGTCGAGTGAAGTCGACACCGCGCGAGATGCTCACTGGAACAACGAGGGGCGCGTCCG 4015  
Db 7626 GACCCGAGATGATTCGAGACGCGCGAGATGCTTCCGCGGAGAACCGCGGACCCGCGTGG 7567  
QY 4016 GAGGCGAGACACCGTTCCGCGAGCACTCGATCTGTGTCAGCGCGAGCGGTGCTCCGT 4075  
Db 7566 CAGGCGAGATGACCTGCTCCGCGAGGTGTTGCTGTCTCCGAGGAGAACCGCGGTCTCGGT 7507  
QY 4076 GCGGTAGGGGCTCATTTGGTGCAGGCGGCGGAGAGATGAGGCTGTGCGACGACTTCTC 4135  
Db 7506 GCGGTAGGGGCTCCGCTGTGTGCTATCCGCGGAGAGATAGGCTGCGCGACGACGTTCTC 7447  
QY 4136 CCGCGCGAGAGATGCGCTGACCGCGGTCAAGCTGATGCGGTGCGGTGCGGTCCGACG 4195  
Db 7446 CCGCGCGAGAGAGATGACGAGAGGTCGCGGTGAGTCCGGTGACGCGCGCTCGATCAG 7387  
QY 4196 GATGATCAGTACTGTGATGTTGCTCTCTGTCCGCGCGGAGCGAGTGCACGCTGACGCGC 4255  
Db 7386 GCGGATGAGTATGATGATGTTGTTGCGTCCGCGGTGTGACTGATGACCGGACGCGC 7327  
QY 4256 GCGTACGTGCGCGAGCTGCTGTGTGATGAGCGCGGTGTGACCCGCGTGTGCTCCCTGCT 4315  
Db 7326 GGGGAGGCGCGAGAGTTCCTGCGCGGTATGAGCTGTATGAGCGCGGTGTGCTGCGACCG 7267  
QY 4316 CTCGCGGAAACGCGTCAAGGAGGTGAGCCCATGCGCGCGCGCACTCGCTCATCTTGGC 4375  
Db 7266 GTCGCGGAAACGCTCCAGGAGGTGAGGCGCGCGCGCGCTCGCTCATCTTGGC 7207  
QY 4376 GTTGGGCGCGATCTGCGTACCACTTGTCCGCGCGAGATGCGAGTGTGATGCTGCGCGC 4435  
Db 7206 GTTATGATC-----CGCGCGCGAGCGCGCTGTGAGCGCGCGAGCGCGCGCG 7153  
QY 4436 GATCCGTTGCGCGAGCGCGCGCTGTGTGTGACACCGCGCGCGCTCGACAGCGCGTGCAC 4495  
Db 7152 CAGCGCGGTGCGCGCGGT 7093  
QY 4496 CCGCTTGT 4555  
Db 7092 GACTTGT 7033  
QY 4556 ACCCGT 4615  
Db 7032 CCGT 6973  
QY 4616 GCGCATCTTGT 4675  
Db 6972 GCGCATGT 6913  
QY 4676 CCGCGCGGT 4735  
Db 6912 CCGCGCGGT 6853  
QY 4736 CCGGT 4795  
Db 6852 CCGGT 6793  
QY 4796 CCGGAGGTCTATGTGACGCGATGATCACTTCCGCGGTGACGCGCGCGCGCGCGCGAG 4855  
Db 6792 GACGAGAGGTCTATGTGACGCGATGATCACTTCCGCGGTGACGCGCGCGCGCGCGAG 6733  
QY 4856 TTCCAGAGGCGAGCGGT 4915  
Db 6732 TTGAGGTCCCGCGGT 6673

QY 4916 GACCCGCGCGCTGCAATCCCGGACGAGGGGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGT 4975  
Db 6672 GATGCGCTGTGTGATCTCCCGAGACGAGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGT 6613  
QY 4976 GCGCCATCTCCAGCGCGCGCGAGAACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5035  
Db 6612 GCGCCATCTCCAGCGCGGT 6553  
QY 5036 CAGGGGT 5095  
Db 6552 GAGGGGT 6493  
QY 5096 CAGCGCGT 5155  
Db 6492 CATGCTGTCTTCTCCCGGT 6433  
QY 5156 TTGACGT 5215  
Db 6432 CCGAGGCGCGAGCGCGCATGACGAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5215  
QY 5216 TTGCGGGGACCGCGCTGT 5247  
Db 6372 CCGAGGACGCTGT 6341

RESULT 15  
AF055579/c  
LOCUS  
DEFINITION  
AF055579 Streptomyces antibioticus putative 3-ketoreductase (oleW),  
2,3-dehydratase (oleV), drp-4-keo-6-deoxyglucose 3,5-epimerase  
(oleU), drp-D-glucose synthase (oleS), dehydratase (oleB),  
4-ketoreductase (oleU), dehydratase (oleU), reductase (oleT),  
glycosyltransferase (oleI), aminotransferase (oleN2), and  
glycosidase (oleI) genes, complete cds.  
AF055579  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Streptomyces antibioticus  
Streptomyces antibioticus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (bases 9806 to 15052)  
Qutros,I.M., Aguirrezabalaga,I., Olano,C., Mendez,C. and Salas,J.A.  
Two glycosyltransferases and a glycosidase are involved in  
oleandomycin modification during its biosynthesis by Streptomyces  
antibioticus  
Mol. Microbiol. 28 (6), 1177-1185 (1998)  
9680207  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
2 (bases 1 to 9805)  
Aguirrezabalaga,I., Olano,C., Allende,N., Rodriguez,I., Brana,A.F.,  
Mendez,C. and Salas,J.A.  
Identification and expression of genes involved in biosynthesis of  
L-oleandrose and its intermediate L-olivose in the oleandomycin  
producer Streptomyces antibioticus  
Antimicrob. Agents Chemother. 44 (5), 1266-1275 (2000)  
10770761  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
3 (bases 9806 to 15052)  
Qutros,I.M., Aguirrezabalaga,I., Olano,C., Mendez,C. and Salas,J.A.  
Direct Submission  
Submitted (25-MAR-1998) Biologia Funcional, Universidad de Oviedo,  
Julian Claveria s/n, Oviedo, Asturias 33006, Spain  
4 (bases 1 to 15052)  
Qutros,I.M., Aguirrezabalaga,I., Olano,C., Mendez,C. and Salas,J.A.  
Direct Submission  
Submitted (13-SEP-1999) Departamento de Biología Funcional e  
Instituto Universitario de Oncología de Asturias (I.U.O.P.A.),  
Universidad de Oviedo, Oviedo 33006, Spain  
REMARK  
COMMENT  
FEATURES  
On Sep 16, 1999 this sequence version replaced gi:3023163.  
Location/Qualifiers  
1..15052  
/organism="Streptomyces antibioticus"  
/mol\_type="genomic DNA"



gene  
CDS  
/ecrain="ATCC.11891"  
/db\_xref="ATCC:11891"  
/db\_xref="taxon:11890"  
complement (590. .1576)  
/gene="olew"  
complement (590. .1576)  
/gene="olew"  
/function="possibly involved in the C-2 deoxygenation step  
in the dtdp-L-oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="putative 3-ketoreductase"  
/protein\_id="A055450.1"  
/db\_xref="GI:5902166"  
/translation="MPSRLRFGVLAADIALRTPTALAHPOVTVAVSRDTARA  
ARPAARCCAVRSHQALLDDDDIDALYPLRVVWHPVWEALIRGRHAKRPLTA  
TRSGAEDLIALMRGVLIMENFTSLHNAOHGYTDLIRGTIGELRSLSAATTIPK  
PEGDIRYQDPVGGALLDIGIYPLRALHFLRGLDPLAAGAVALRRERRNVVSGHLL  
TTPHGVVAELAFQMEHAYRBSBYTLFTAGRLRLDRAAPETTRPVEIHRDADLIV  
DLPPDAQFANLVRDVFVAVRSGPRLTQGHADAVRQADLVERVMAVARVWC"  
complement (1596. .3020)  
/gene="olev"  
complement (1596. .3020)  
/gene="olev"  
/function="involved in the C-2 deoxygenation step in the  
dtdp-L-oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="2,3-dehydratase"  
/protein\_id="A055451.1"  
/db\_xref="GI:5902167"  
/translation="MIMGIPAMSGVPTPAGSSTCAFLSMDARRRRLTYE  
LVPRRLSGMOFDENTGNLHTSGRFPISIGLARVTDHCFWSTOPIIYQPELIGL  
LHYVRFGIHLVYQAKNERNIGLQLSPTVQATRSNYTRVHGSGVRLYEASR  
GRGVLDVLSQSGSFLHGRNRMVVEADDVLDLDDFMWISLGGKRLMLPHLV  
NMDRTVLSCLPPDPADGRQPPAPAFPAAYRSLIRGATLHTGELLWLTDR  
SRRLVOORVLEETAFSGWKRDDHAIAHXOGDYFRITGVSVRASREVSWQPLA  
PVGSLAFTYRIRGVLYHLHARTAGLNGEMAPVOCPLNRYRAVAEYRPAY  
LDVYLSDPGRIRDTLOSEEGRFHIAENRYVVAEDDEPVEVPDPMLTLHQL  
ALLHSNRYVRAISLVACTOALS"  
3356. .3970  
/gene="oleL"  
3356. .3970  
/gene="oleL"  
/function="epimerization at carbons 3 and 5 in the  
biosynthesis of L-oleandrose"  
/codon\_start=1  
/transl\_table=1  
/product="dtdp-4-keto-6-deoxyglucose 3,5-epimerase"  
/protein\_id="A055452.1"  
/db\_xref="GI:5902168"  
/translation="MELLIDVNGAMLYTPEIMRDRGRTLEFRGRTQEKIHPISLA  
QANCSVRKAPCAASTFTPPQAKTYTCASGVLDVVDVRGSGTFEGMAVLD  
AARQOGLYLAEGLAHAFMALTDATVYVLCSPYVAEERAVDPLDPAIGLEWPTD  
IVPVGEPTTRPWRPRRPGILPDYEGVPGALHRGGRGRTGP"  
3967. .5037  
/gene="oleS"  
3967. .5037  
/gene="oleS"  
/function="activation of glucose-1-phosphate into  
dtdp-D-glucose"  
/codon\_start=1  
/transl\_table=1  
/product="dtdp-D-glucose synthase"  
/protein\_id="A055453.1"  
/db\_xref="GI:5902169"  
/translation="MKALVLAGSGGTRLRPTHTSAKQLVAVANKPVLYGLAIAAA  
GITVGLVGTAGVPRVAGGAKFGDIDITYRQSRPLGAHVLTAHTVLGDDVY  
MYLDGNPTVGGTIDLVRRPRGRRAAIIITHTYSDSGVAVLDDGGVYGEED  
RHPKSDLVGVYFTTPAIEHVAIRIESMGELEITTHAILDNGADIOSMTKEG  
WKDTGNVADMLKVRKTVLELEPRIEGTVDHTVTGIVVVGSGARVNSIMKPAII  
GAGPEISDSTYGFTSGVGNCRITGSEWEPFIMLASAITVRRIEGLIGRNVVTO  
SLHAPNARFVLGDHSKVEIQS"

gene  
CDS  
5034. .6041  
/gene="oleB"  
5034. .6041  
/gene="oleB"  
/function="4,6-dehydration during 6-deoxysugars  
biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="dehydratase"  
/protein\_id="A055454.1"  
/db\_xref="GI:5902170"  
/translation="MWLVTTGAAGFTGSRVYHLLHENTRGRPRAPITYLDTKTYAG  
VIGNVDPDPAYTVTFGRGLADAPLVDSLMADADVHFAETHYDRKSTSGITKTVN  
LGTQVLLDALRHGVGPFWYVSTDEYVGSIEHGSWPEHOPLCNPSYSAKASDLA  
LSYHRTGDLVRYTRCSNNYGPPEKTYPLFTVTLIDLRPLVLDGGLNREMLYH  
DDHCLVDVLRTOGRGVEYHIGGTELTNRDITGLLDAFGVMVDVDRVADRKGD  
RRYALDCAARAADBLGTRPRDPABJARTIDWIRDRAMWEPKAKRPAAPAPRSGG  
P"  
6038. .6925  
/gene="oleU"  
6038. .6925  
/gene="oleU"  
/function="involved in L-oleandrose biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="4-ketoreductase"  
/protein\_id="A055455.1"  
/db\_xref="GI:5902171"  
/translation="MRWLTGAAGMTGREIVRELANEEDVALGHDHLDVTPSAVR  
AALAEHPGIVAVCAATVAVDDABETBAAALIMAEAPRLAAGLRPHRHGVHST  
DYVPGDARTVADHDPTAPRSAYGRTKRGEAVLITATVAVLWVLTGKTSF  
VRYMIERARAGGAIIDVADRGQPTVGTADADIIVAGRRPVGHILAHATNASATVY  
DLAOFERLIDRPGRVPRPTGAFFRAPRAPVSYGHDNRWGTGIALPDRWSALR  
EARPDLIAAHPTTRGAA"  
6922. .8103  
/gene="oleNT"  
6922. .8103  
/gene="oleNT"  
/function="possibly involved in the transamination,  
together with olen2, during the biosynthesis of  
D-desosamine"  
/codon\_start=1  
/transl\_table=1  
/product="dehydratase"  
/protein\_id="A055456.1"  
/db\_xref="GI:5902172"  
/translation="MKRGVLDLALFGDAAFLQPYMGPRNTGDRRLRLLEVALN  
RWLNGSPVLRFEFORIADLAGVRNVCATONATAGLQLIRBAEYTGVIYMPSTVA  
TAHAVRWLGVRPVFCIDIDPTGCLDPLVEAAVTPTGAILGVHLMGRPSRVDLAAI  
AAEHGLKLFDAHAALGCTSRORVNSFGDAEFSFATATVNSFEGGGVTDODTBA  
ERLRALNFGIGHDGVAGINAKMSEAAAGLTSLEAFADAVASNAAYELYQELS  
GLGGVRLIDYDPAERNNYVYVLLIDGVYGLHRDLITLLRAENVVAOYFSGCHQ  
REPRTRHPVSLHTHELAQVIALPTGPAVSREDIRVCDIIRVAAHHPRTAQAQ  
A"  
8225. .9682  
/gene="oleT"  
8225. .9682  
/gene="oleT"  
/function="possibly responsible for the 3,4-reduction step  
in D-desosamine biosynthesis"  
/codon\_start=1  
/transl\_table=1  
/product="reductase"  
/protein\_id="A055457.1"  
/db\_xref="GI:5902173"  
/translation="MALQVDRTPIDPIAVCAMPGNSAAAAALPILIEDAPGTGLEP  
DRLAENHILARRGYTEPFPILSARGGLGIDRTAFVLAHRTPLAAVARGGA  
GMAWHTIILPERSGVDAVARGEPAPVYPPGSCMFRCHFCRVAVGAFQOQA  
LITQNAMFASLIDBMPDNPHAYLSGGLPLTWGQDLYVRAAAGFKLSITYNF  
ALTRQCTIDRQGLMDIYALRTSLYGEDVVAATTKKGAFORKDLTFQOALRRR  
RUPRVGLANTITLPRGARGLTGADYTGDLNDGPPDPVPFUTLREYSGRPOKAP  
BERVELEHGAFAEERLRTAPSLHVDYGYALOSLRGVDAELPRILPETERPTAHFQ  
VAVQVDLGDVLYIREAGFPGLQGAERYVAGRLTTGTLEBEVVARFVTEGRQVAPRG

gene  
CDS  
EXFELDFDQTVTARINQMETIADGNABHRFLR"  
957. .11231  
/gene="ole1"  
957. .11231  
/gene="ole1"  
/function="intracellular inactivation of oleandomycin by  
glycosylation"  
/codon\_start=1

Query Match 14.2%; Score 1269.8; DB 1; Length 15052;  
Best Local Similarity 66.6%; Pred. No. 3.9e-82;  
Matches 1940; Conservative 0; Mismatches 907; Indels 66; Gaps 6;

QY 2395 GCGGATCAGGCGCCGATGCTGATCCACCCGTCGACAGTCCCGTTGAGTTGTTGAG 2454  
DB 9679 GCGCAGGAAACCCCGGCTCTCCGCGCAGCCGTCGCGATTCGCTCTCACTGCTTGA 9620  
QY 2455 CCGGCGGTCACGCACTGCTGAAACGCTGAGAGAAACTCGTCCCGCGCTGCGATC 2514  
DB 9619 ACGTGCCTGACGCTCTGCTGCAAGCGCTGAGAAAGTACTCTTCGCGCGGCGCGCGC 9560  
QY 2515 GATGCTGCGGCGCGACGTGACGAAGTGTGAGACGACGAGTCAAGCTCCGCTCGGAGT 2574  
DB 9559 CACTTCGCGGCGCTCGGTCGAAACGCGCGACGACTCTCTCAAGTTCGCTGCGGAGT 9500  
QY 2575 GACCGGCGGCGACGATGAGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2634  
DB 9499 CACCGGCGGCG 9440  
QY 2635 CAGGTAGACGTCTCCGACGAGTGTGATCTGACCGCGACCTGCGAGTGGCGGTCGCG 2694  
DB 9439 GAGGTAGACGTCTCCGACGAGTGTGATCTGACCGCGACCTGCGAGTGGCGGTCGCG 9380  
QY 2695 CATGCTTTCCGCGCGGATTCGCGACAGCTGAGCGGTGACACCCGACGAGGCTCTCA 2754  
DB 9379 CATGCTCTCCGCGCGGATCTCTCGGCAAGTTCGCGGTGACAGCGCGCGCGCGCG 9320  
QY 2755 CCGGTACCCAGGTCGCTGTCATGCGCGGCGGTCGCGTGGCGGCGGATGACGAAACG 2814  
DB 9319 GCGGTAGCGGTATGACGTGACGAGGACCGGCGCGGTGCTGATCCGTTCTCGAAACG 9260  
QY 2815 GCGGAGGCGCTCCCGCACTCGGTCCGTTCCGCTCGACCAACTCGCGCTCTCGCGAC 2874  
DB 9259 GCGGAGGCGGTCCTCAAGTTCCAGCGTTCTCGGGGCGAGTTCCGCTCGGCGCGTCC 9200  
QY 2875 GCTGTAGTCTTCGCGGACGCTGACGAAGTGCAGCGCGCGGTCGCGGCTGACCTCTT 2934  
DB 9199 GCTGTAGTCTTCGCGGAGGCTGAGGAATGTCACGCGAGCGGTCCGCGCGCTGAG 9140  
QY 2935 TTCGCGGATGAGTGCACGAGTGCATGAGCGGTTGCGCGCGCGCGCGCGCGCGCG 2994  
DB 9139 GTGCGGATGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9080  
QY 2995 GTTGAAGCGAGCGCGACCGGTGCGCGCTTCGCGCGCGCATCCGACGAGAGTCCCGAG 3054  
DB 9079 GTTCAAGCGCGACGCGCGCTTCGAGCTTCCCGCGCGAGGCTTGAAGCGGCTGAG 9020  
QY 3055 GTTCTTCTGACCGGTTCAAGCGCGCGCTTCGCGGTGCGGTGAGCTGTGATCTGCTT 3114  
DB 9019 GTTGTCTTGAAGCGCGCTGAGAGCGCGCTTCTTCGGTGCTGTCGACGATGCTGCTC 8960  
QY 3115 GTTCAAGCGCGTACAGAGGATGACGACCGCGCGCGAGTCCACAGCGCGCGCGCGCG 3174  
DB 8959 GCTAGAGCGCGTACAGAGGATGACGAGCGCGCGTACAGTCCACAGCGCGCGCGCGCG 8900  
QY 3175 CAGGCTGTGCTGAGTGAAGCGCGGCTGCTGATGACGCTCAAGCGCGAGCGCGCGCG 3234  
DB 8899 GAGGCTGTGCTGAGGCGCGGATGAGTGTGATGAGGAGAGTGTGAACCGCGCGCG 8840  
QY 3235 GCGCGCGTGCACACACGCGCGGTGATCCCGGCTTGTCAACGCTTCAAGCGCGCGAG 3294  
DB 8839 CCGCGCGACGCGGACGAGGTGCGCGGCTGCGCGGCTTGTCAAGCGCGTCAAGCGCA 8780

QY 3295 GAGGTATCATGCGCGGTGGGTTCTCCAGCGGACCTCTGTCATGACCGAGGTGAGATCTC 3354  
DB 8779 GAGGTATCAGGCGCGTGGGGTTTGTCTGTTGAGCATTCGCTGATACGAGGACCGAATCC 8720  
QY 3355 GTTTCGGGCGACACACGACTCCGCGTGTGATCTTGGACCTGTACCCGTAACGAGATG 3414  
DB 8719 GTTTCGCTCGTACAGGCGCGACTGCTGATGAGCGGCGCGGTCACCGGTATCCGAGAA 8660  
QY 3415 ACAGCGGAACTATGACAGTGGCGCGGCTGATGAGCGCGCGCGCGCGCGCGCGCG 3474  
DB 8659 GCGAGCGAACTATGACAGTGGCGCGGCTGATGAGCGCGCGCGCGCGCGCGCGCG 8600  
QY 3475 CCGCGTACCGGACGCTTCAACGCGCGCGCGGCTTCAAGCGCGCTGATGCTGCTCCA 3534  
DB 8599 GCGCTTACCGCGCGCTTCAAGGACGCGCGCGCGCTTCAAGCGCGAGATGCTGCTCA 8540  
QY 3535 GTACTTCCCGCGCGCGCGCTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3594  
DB 8539 GTACATGCGCGCGCGCGCGCGCTTCAACGCGCGCGCGCGCGCGCGCGCGCGCG 8480  
QY 3595 GTTACAGACTGCGCGAACCGCGCGCGCTTCAAGTGTGAACCGAGCGCGACTTCTCA 3654  
DB 8479 GCGAGAGAGCGCGCGCGAGTGGCGGCTTCAAGGCGCGCGCGCGCGCGCGCGCG 8420  
QY 3655 CCGGCGTGAACCGCGAGGCGCGCGTGAAGGCGG-----TGAAGTTCCGCGAGCG 3702  
DB 8419 GCGCGTGAAGGCTTCTGTCGTCGTAAGTGGCGCGAGGCGTGAAGTGTCTCGCGAGCG 8360  
QY 3703 GACCGCTGCTGCTTCCGCGCGGTCGATGCGGACCGGTAAGGCGGCTGACCTCGCGCT 3762  
DB 8359 GTTCCGCTTCAAGTCCGCGTCCCGGCGCGCTCTCTGATGAGGCGCGCGCGCGCG 8300  
QY 3763 GACCGCGCGACCG 3792  
DB 8299 GCGCGCGCTGTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8240  
QY 3793 GTTCCGCTGCG 3846  
DB 8239 GACCTGAAATGCGACCGCGTCCGCTCTTGTCTTCTTCTTCTTCTTCTTCTTCTT 8180  
QY 3847 -GCTTCTCATGACGTGCGAACCTTCTGCGCTGCTGCTGCT-----GCGAGATCA 3897  
DB 8179 GCGTCTCCTTCCGCGGAGGTCGCGCTCTCTGCGCTTCCCGCGCGCGCGCGCGCG 8120  
QY 3898 AACCGTTCGCTTCCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3954  
DB 8119 GCGCTTCTGACCTGCGGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8060  
QY 3955 CGAGTCCGATGACGTGCGACACCGCGCGGATGCTCTCATCTGACACCGAGGCGCG 4014  
DB 8059 CGACCGGATGATGTCGACGACGCGTCCGATGCTCTCGCGGAGACCGCGGAGAC 8000  
QY 4015 GAGAGCGGACGACCGGTTCCGCGGACCACTCGGCTGCTGTCAGCGCGCGCGGCTCG 4074  
DB 7999 GAGAGCGGATGACCTGCTCGCGGAGGTGCTCGGTCGCGGAGACCGCGGAGTCTCG 7940  
QY 4075 TCGGATGAGGCGCTGATCTGATGAGCGCGGAGAGAGTGAAGGCTGTGCGACCTTCT 4134  
DB 7939 TCGGATGAGGCGCTGCGGCTGATGAGCGCGGAGAGAGTGAAGGCTGTGCGACCTTCT 7880  
QY 4135 CCGCGCGGAGATTCGCTGCGCGCGGTCACCGGTGCAATGCGGTCGCGGCTCTCA 4194  
DB 7879 CCGCGCGGAGAGGTCGACGAGAGGTCCCGGTGAGTCCGCGGTCGCGGCGGTGATTA 7820  
QY 4195 GAGTATCACTGATGATGCTCTCTCTCGTCCGCGCGGAGCGAGTCAAGTCAAGC 4254  
DB 7819 GCGGCGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7760  
QY 4255 GCGGTACGTCCGCGCGCTGCTGATGAGCGCGGTGATGATGATGATGATGATGATGAT 4314  
DB 7759 CCGGAGGCGCGGAGGTTCTGTCGCGGTGAGCTGATGATGATGATGATGATGATGAT 7700  
QY 4315 TCTCGGCGAAGCGCGTGAAGGAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 4374

Db 7699 CTTGGGCGAAGCGCTCCAGGAGTCAAGCCCAATGGCGGCGCGCTCCGTCATCTTGG 7640  
QY 4375 CTTTTGGCCGATCTGGGTGACCACTTGTCCGGGCGATGCCAAGTTGTGATGGCCC 4434  
Db 7639 CTTTGAATG-----CGGCGCCGACCGCTGTGTGGCCAGGCCGAAATTGTGACGCGC 7586  
QY 4435 GATTCGTTTCGGCCAGCAGGCGCGTGTGTGACCAACCGCCCGCCCTCGAAGCGGTGA 4494  
Db 7585 GCAGGCGCTGCGCGCGGT 7526  
QY 4495 CCGCTTTGTGGCGGTGAAGCTGAACACTTCGGCGTCAACCGATCCGCCACCGAACGTC 4554  
Db 7525 CGACTTGTGGCGGTGAAGCTGAAGACTTCGGCGTCCGAAAGAACCGAGCGCGCT 7466  
QY 4555 CACCGCTGTGCAAGCCGAGGCGGTGGCGCGGTGCAAGAGAGCTTGAACCTGGTGTGG 4614  
Db 7465 GCGGTGAGTGCATCCAGAGGCGGTGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 7406  
QY 4615 CCGGATCTTCGCGAGCGCTCCACAGGTGTGGTCCGCGCCACAGGTGTACGCCGACGA 4674  
Db 7405 CCGCTATGGCGCGAGTTCTGTCAACCGCGCTGGCGCCGCCAGAGGTGACGCCAGGA 7346  
QY 4675 TCGCGCGGTCTTGCGGGTGAAGAGCGCGCGAGCTGTCCGGGTCAACAGACCGGTG 4734  
Db 7345 TCGCGCGGTGGCGGGGTGAAGCGCGCTTCAGAGTTGGGGTCAAGGCAAGCGGTGT 7286  
QY 4735 CCGGTGCAAGTCCGCAAAACACCGGTGTGATCCGACAGCTGCGCGGTGGCGGGTGG 4794  
Db 7285 CCGGTGATGTTCGCAAGAGCGGGCGAGTCCGAGCCACCGGACGCGGTGGCGGGTGG 7226  
QY 4795 CCGGAAAGTCAATTGACGGCATGATCACTTCAACCGGTGACGTCAACCGGCCGACAGCA 4854  
Db 7225 CGAGAAAGTATCGACGCGCATGATCACTTCCCGGTGACCTCCGCTTCGGCGAGCA 7166  
QY 4855 GTTCAGGGCGACGCTGCGGTGCGATGCGATGCGATGCCCGACCGAGTCCG 4914  
Db 7165 GTTGAAGTCCCGGTGGCGTTGACAGGTGGCGAGCGAGTTCGTACACCGGCCAGTCCG 7106  
QY 4915 CGACCCGGGCTCGAATCTCCCGCACAGGGGCGCGGTGTGTGAGCCAGTTGTTC 4974  
Db 7105 CGATGCGCTGTCAAACTCCCGACAGCGGGCGCGGTGTCAAGCCACCGTTGTCCA 7046  
QY 4975 GGGCCCACTCCAGGCGGGCGAGAACCGCTGCGCGATCGTCCGCGCGGCCAGT 5034  
Db 7045 GGGCCCACTTCAGGCGGTTCAGAGGCGTTTCGGTCAACCGGTGTGGGGCGCCCATGT 6986  
QY 5035 GCAGGGGTGACGAAACGCTTCGGGACCGCGAAGATCCGCAATCGGTGGTACGCGT 5094  
Db 6985 AGAGGGGTGAGAGAGCGCGGTGCGCGCGAAGAGGGCCAGGTCTGTGACGCTCGT 6926  
QY 5095 TCAGCGCGTCCCGGT 5154  
Db 6925 TCATGTCTCTCTCGCGCGGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6866  
QY 5155 GTTGAAGTGTGCTGTGCGAGGATTCGAGTTCGAGTGTGAGGATTCATCCAGCGT 5214  
Db 6865 GCGCAGGCGCGAGCGCGCATGACGAGAGGTGTCCAGGCGGTCCCGCGCACGCGTCTG 6806  
QY 5215 GTCGGGGACCGCGGTGTGAAATCCGCGG 5247  
Db 6805 GCCGAGACGCTGTAGCGGGGCGGGCGGG 6773

Search completed: April 7, 2006, 16:58:07  
Job time : 43636.7 secs

```

TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:
NAME/KEY: CDS
LOCATION: (1) ... (144)
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (928) ... (2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072) ... (3382)
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCII homolog),
OTHER INFORMATION: TDP-megosamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462) ... (4634)
OTHER INFORMATION: megG (megT), mycarosyl acyltransferase, mycarose O-acyltransferase
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651) ... (5775)
OTHER INFORMATION: megDI, deoxysugar transaminase (eryCI, DnrI homolog),
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase,
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS

```

LOCATION: (5822)...(6595)  
OTHER INFORMATION: megDII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (6592)...(7197)  
OTHER INFORMATION: megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dmuv hc  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (7220)...(8206)  
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dmuv homolog),  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (8228)...(9220)  
OTHER INFORMATION: megBII-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (9226)...(10479)  
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;  
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (10483)...(11424)  
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (12181)...(22821)  
OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (12505)...(13470)  
OTHER INFORMATION: megAI, AT-L  
NAME/KEY: misc feature  
LOCATION: (13576)...(13791)  
OTHER INFORMATION: megAI, ACP-L  
NAME/KEY: misc feature  
LOCATION: (13849)...(15126)  
OTHER INFORMATION: megAI, KS1  
NAME/KEY: misc feature  
LOCATION: (15427)...(16476)  
OTHER INFORMATION: megAI, AT1  
NAME/KEY: misc feature  
LOCATION: (17155)...(17694)  
OTHER INFORMATION: megAI, KR1  
NAME/KEY: misc feature  
LOCATION: (17947)...(18207)  
OTHER INFORMATION: megAI, ACP1  
NAME/KEY: misc feature  
LOCATION: (18268)...(19548)  
OTHER INFORMATION: megAI, KS2  
NAME/KEY: misc feature  
LOCATION: (19876)...(20910)  
OTHER INFORMATION: megAI, AT2  
NAME/KEY: misc feature  
LOCATION: (21517)...(22053)  
OTHER INFORMATION: megAI, KR2  
NAME/KEY: misc feature  
LOCATION: (22318)...(22575)  
OTHER INFORMATION: megAI, ACP2  
NAME/KEY: CDS  
LOCATION: (22667)...(33555)  
OTHER INFORMATION: megAI, SEQ ID NO: 14= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (22957)...(24237)  
OTHER INFORMATION: megAI, KS3  
NAME/KEY: misc feature  
LOCATION: (24514)...(25581)  
OTHER INFORMATION: megAI, AT3  
NAME/KEY: misc feature  
LOCATION: (26230)...(26733)  
OTHER INFORMATION: megAI, KR3 (inactive)  
NAME/KEY: misc feature  
LOCATION: (26958)...(27258)

OTHER INFORMATION: megAI, ACP3  
NAME/KEY: misc feature  
LOCATION: (27353)...(28590)  
OTHER INFORMATION: megAI, KS4  
NAME/KEY: misc feature  
LOCATION: (28897)...(29331)  
OTHER INFORMATION: megAI, AT4  
NAME/KEY: misc feature  
LOCATION: (29953)...(30477)  
OTHER INFORMATION: megAI, DH4  
NAME/KEY: misc feature  
LOCATION: (31396)...(32244)  
OTHER INFORMATION: megAI, ER4  
NAME/KEY: misc feature  
LOCATION: (32257)...(32799)  
OTHER INFORMATION: megAI, KR4  
NAME/KEY: misc feature  
LOCATION: (33052)...(33312)  
OTHER INFORMATION: megAI, ACP4  
NAME/KEY: CDS  
LOCATION: (33666)...(43271)  
OTHER INFORMATION: megAI, SEQ ID NO: 15= translated amino acid sequence  
NAME/KEY: misc feature  
LOCATION: (33780)...(35027)  
OTHER INFORMATION: megAI, KS5  
NAME/KEY: misc feature  
LOCATION: (35385)...(36419)  
OTHER INFORMATION: megAI, AT5  
NAME/KEY: misc feature  
LOCATION: (37068)...(37604)  
OTHER INFORMATION: megAI, KR5  
NAME/KEY: misc feature  
LOCATION: (37860)...(38120)  
OTHER INFORMATION: megAI, ACP5  
NAME/KEY: misc feature  
LOCATION: (38187)...(39470)  
OTHER INFORMATION: megAI, KS6  
NAME/KEY: misc feature  
LOCATION: (39795)...(40811)  
OTHER INFORMATION: megAI, AT6  
NAME/KEY: misc feature  
LOCATION: (41406)...(41936)  
OTHER INFORMATION: megAI, KR6  
NAME/KEY: misc feature  
LOCATION: (42168)...(42425)  
OTHER INFORMATION: megAI, ACP6  
NAME/KEY: misc feature  
LOCATION: (42585)...(43271)  
OTHER INFORMATION: megAI, TR  
NAME/KEY: CDS  
LOCATION: (43268)...(44344)  
OTHER INFORMATION: megII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;  
OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (44355)...(45623)  
OTHER INFORMATION: megCII, desosaminyl transferase, desosamine glycosyltransferase  
OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (45620)...(46591)  
OTHER INFORMATION: megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,  
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;  
OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (46660)...(47403)  
OTHER INFORMATION: megH, TEII, SEQ ID NO: 19= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (47411)...(47980)  
OTHER INFORMATION: megH, TEII, SEQ ID NO: 19= translated amino acid sequence  
NAME/KEY: CDS  
LOCATION: (47411)...(47980)  
OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequ  
US-09-679-279-1

Query Match 100.0%; Score 600; DB 3; Length 47981;  
Best Local Similarity 100.0%; Pred. No. 1.3e-102; Indels 0; Gaps 0;  
Matches 600; Conservative 0; Mismatches 0;

QY 1 CCGCGCTCGCCGAGGCTACACCCGCGGGGTGAGGTGACTGCGCTACCGCAGTGGTG 60  
Db 36307 CCGCGCTCGCCGAGGCTACACCCGCGGGGTGAGGTGACTGCGCTACCGCAGTGGTG 36366  
QY 61 AGGAGAGCCCGGTGACCTCCCGGTCTACCCGTTCCAGACAGAACTTCTGGCTCCCG 120  
Db 36367 AGGAGAGCCCGGTGACCTCCCGGTCTACCCGTTCCAGACAGAACTTCTGGCTCCCG 36426  
QY 121 TCCCGCTGGGCGGGTCCCGGACACCGGCGACAGGTGGCGTTACAGCTGGCTGGGACC 180  
Db 36427 TCCCGCTGGGCGGGTCCCGGACACCGGCGACAGGTGGCGTTACAGCTGGCTGGGACC 36486  
QY 181 CCGTCCGACTCGGCGGGTCTCCCTGGCGCGAGCGGGTCTGTGTGTGACCGGAGCGGACG 240  
Db 36487 CCGTCCGACTCGGCGGGTCTCCCTGGCGCGAGCGGGTCTGTGTGTGACCGGAGCGGACG 36546  
QY 241 TACCCCGGCTTGACGAGCAGTGGTCCGACCGGCTGAAACAGCGCGGGGCGACCGCTCG 300  
Db 36547 TACCCCGGCTTGACGAGCAGTGGTCCGACCGGCTGAAACAGCGCGGGGCGACCGCTCG 36606  
QY 301 TGTGTGACACCGGCGAGTCCCGCGCGCGGATCCGCGCGCACTCGACGCGCTGACGCGCA 360  
Db 36607 TGTGTGACACCGGCGAGTCCCGCGCGCGGATCCGCGCGCACTCGACGCGCTGACGCGCA 36666  
QY 361 CCGCGCTGTCACGTGTGTCTCTGTGTGCGGCTCGCGCGGCGGGTGTGTGACGAGACC 420  
Db 36667 CCGCGCTGTCACGTGTGTCTCTGTGTGCGGCTCGCGCGGCGGGTGTGTGACGAGACC 36726  
QY 421 CCAGCTTGACACCTTCGCGTGTGTGTTCAGGCGCTCGCGCGAGCGGAGTTCAGCTCCCGC 480  
Db 36727 CCAGCTTGACACCTTCGCGTGTGTGTTCAGGCGCTCGCGCGAGCGGAGTTCAGCTCCCGC 36786  
QY 481 TGTGTGTGTGTGACACGAGGAGCGCGCGCGCTGTGACCGTCCGAGACGAGCTGTGATCCGCGCC 540  
Db 36787 TGTGTGTGTGTGACACGAGGAGCGCGCGCGCTGTGACCGTCCGAGACGAGCTGTGATCCGCGCC 36846  
QY 541 AGGCGATGGTGGTGGGCTCGCGCGGGTGTGTGCGGTGAGTCCCGCGCGCGGGTGGTGG 600  
Db 36847 AGGCGATGGTGGTGGGCTCGCGCGGGTGTGTGCGGTGAGTCCCGCGCGGGTGGTGG 36906

RESULT 2  
US-07-642-734C-3.

Sequence 3, Application US/07642734C  
Patent No. 5824513  
GENERAL INFORMATION:  
APPLICANT: Katz, L  
APPLICANT: Donadio, S  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
TITLE OF INVENTION: Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Edward H. Gorman  
STREET: Abbott Laboratories D377/ABED-2 One Abbott  
CITY: Park Rd  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckert, Andreas M  
REGISTRATION NUMBER: 32652

REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 238  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..10722  
OTHER INFORMATION: /codon start= 19  
OTHER INFORMATION: /function= "gene eryA"  
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19..1470  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 97..1482  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1693..2670  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3406..3921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4171..4428  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..10722  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..5847  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6054..7026  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 7165..9216  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: dehydratase and enoylreductase domains m"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9433..9984  
OTHER INFORMATION: /function= "approximate span  
OTHER INFORMATION: beta-ketoreductase of module 4"



FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10225..10483  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 4"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10723..20235  
OTHER INFORMATION: /codon\_start= 10723  
OTHER INFORMATION: /function= "gene -eryA"  
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6  
OTHER INFORMATION: 6-deoxyerythronolide B formatic"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10723..15165  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10831..12174  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACP synthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 12379..13350  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14062..14610  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14857..15114  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15166..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15172..16569  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACP synthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 16768..17721  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 18379..18921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19149..19398  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19492..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: thioesterase domain of module 6"  
US-07-642-734C-3

Query Match 47.8%; Score 286.6; DB 2; Length 20235;  
Best Local Similarity 69.7%; Pred. No. 1,1e-44;  
Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

Qy 1 CCGCGCTCGCGAGCGGTACCCCGGGGGTGAAGTGCATCGTCCGTACCGCACTGGGCTG 60  
|||  
Db 13304 CCGCGCTCGCGAGCGGTACCGCGAGGCGTGAAGGTGCACTGAGCCCGCTTCCCG 13363  
|||  
Qy 61 AGGAGCGCCCGGTGACCTGCGGCTGACCGGTTCCAGAGACAGAACTTCGAGTCCCG 120  
|||  
Db 13364 ACCGCGCGCCGCTGACCTGCGGCTGACCGGTTCCAGAGAGCGGTTACTGCTGCCA 13423  
|||  
Qy 121 TCCCGCTGGGCG---CGGATCCCGACACCGCGACAGAGTGGGCTTACAGCTCGCTGAC 177  
|||  
Db 13424 TCCCGACCGCGCGCGCGCGACGAGAGACAGACGACTGGCTTACAGGCTATGAGC 13483  
|||  
Qy 178 ACCCGCTGACCTCGGCGGCTCTCTCGGCGCGAGCGGCTCTGCTGCTGACCGAGCG 237  
|||  
Db 13484 GGAAGCGCGAGTGGAGAGCGCTTCTGCGCGCGAGCGGCTGCTGCTGCTGACCGAGCG 13543  
|||  
Qy 238 CAGTACCCCGCGCTGACCGAGAGTGTCCGCGACCGGCTGGAAAGCGGGGCGACCG 297  
|||  
Db 13544 GCGTGCCTGCGAGTTGTGAGACGCGCATCCGAGTGAAGAGAGAGAGCGGTCGACG 13603  
|||  
Qy 298 TCGGTGTGACACCGCGAGTGGCGCGCGGATCGCGCGGCACTGACGCGCTGACG 357  
|||  
Db 13604 TCTGACCTGCGAGTGGAAATCCGTTGACACATCGGCACTGAGGCGCGGACGA 13663  
|||  
Qy 358 GCAACCGCTGTGCACTGTGCTCTCTGCTGCGGCTGCGCGAGGCGGCTGTGCAAG 417  
|||  
Db 13664 CCGACGCTGTGCTCACTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13717  
|||  
Qy 418 ACCCGACCTGACACCTTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477  
|||  
Db 13718 ATCGTGTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13777  
|||  
Qy 478 CCGTGTGCTGAGTGAACAGAGGAGCGCGCGCGCTGACCGTGGAGAGACGACTGATCGG 537  
|||  
Db 13778 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13837  
|||  
Qy 538 CCGAGGCGATGT 597  
|||  
Db 13838 CGAGGCGATGT 13897  
|||  
Qy 598 GTG 600  
|||  
Db 13898 GCG 13900  
|||

RESULT 3  
US-08-439-009A-3  
; Sequence 3, Application US/08439009A  
; Patent No. 6004787  
; GENERAL INFORMATION:  
; APPLICANT: Donadio, S  
; APPLICANT: Katz, L  
; APPLICANT: McAlpine, J B  
; TITLE OF INVENTION: Method of Directing Biosynthesis of  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Steven F. Weinstein  
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/439,009A  
; FILING DATE: 11-MAY-1995  
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Casuto, Dianne  
REGISTRATION NUMBER: 40,943  
REFERENCE/DOCKET NUMBER: 4952.US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 847-938-3137  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 238  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..10722  
OTHER INFORMATION: /codon\_start= 19  
OTHER INFORMATION: /function= "gene eryA"  
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19..4470  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 97..1482  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1693..2670  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3406..3921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4171..4428  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 3"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..10722  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 4471..5847  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6054..7026  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 7165..9216  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: denhydratase and enoylreductase domains m"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9433..9984

OTHER INFORMATION: /function= "approximate span  
OTHER INFORMATION: beta-ketoreductase of module 4"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10225..10483  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 4"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 10723..20235  
OTHER INFORMATION: /codon\_start= 10723  
OTHER INFORMATION: /function= "gene =eryA"  
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6  
OTHER INFORMATION: 6-deoxyerythronolide B formatio"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10723..15165  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10831..12174  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 12379..13350  
OTHER INFORMATION: /function= "approximatr span of  
OTHER INFORMATION: acyltransferase domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14062..14610  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 14857..15114  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 5"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15166..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15172..16569  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 16768..17721  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyltransferase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 18379..18921  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19149..19398  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: acyl carrier domain of module 6"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19492..20235  
OTHER INFORMATION: /function= "approximate span of  
OTHER INFORMATION: thioesterase domain of module 6"  
US-08-439-009A-3

Query Match 47.8%; Score 286.6; DB 3; Length 20235;  
Best Local Similarity 69.7%; Pred. No. 1,1e-44;

Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

```
QY 1 CCGGCTGCGGAGGCGTAAACCCGGGGGTGAGGTGACTGCGCTGACCGCACTGGAGT 60
    |||||
Db 13304 CCGGCTGCGGAGGCGTAAACCCGGGGGTGAGGTGACTGCGCTGACCGCACTGGAGT 13363

QY 61 AGGAGCGCGGAGTGAAGTCTGCGCGGTCTACCGGTTCAGAACAGAACCTTGTGGCTCCCG 120
    |||||
Db 13364 AGGAGCGCGGAGTGAAGTCTGCGCGGTCTACCGGTTCAGAACAGAACCTTGTGGCTCCCG 13423

QY 121 TCCCTCTGGGCG--CGGCTCCCGCAACCGCGAGAGTGGCGTTTACCACTGCGCTGGC 177
    |||||
Db 13424 TCCCTCTGGGCGCGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 13483

QY 178 ACCCGCTGAGCTTGGGCGGTCTCTCTCTGCGCGGAGAGAGAGAGAGAGAGAGAGAG 237
    |||||
Db 13484 AGGAGCGCGGAGTGAAGTCTGCGCGGTCTCTCTCTGCGCGGAGAGAGAGAGAGAGAG 13543

QY 238 CAGTACCCCGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 297
    |||||
Db 13544 GCGTGGCTGCGAGATTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 13603

QY 298 TCGTGTGTGAGACCGCGAGTGGCGCGCGCGCGAGATCGGCGCGCACTCGACCGCTGAG 357
    |||||
Db 13604 TCGTGTGTGAGAGAGTGAATCCGTTTGAACATCGGACCGCACTGAGAGCGCGCGCA 13663

QY 358 GCACCGCGCTGTCACTGTGTGTCTGTCTGCTGCGGCTGCGAGAGAGAGAGAGAGAG 417
    |||||
Db 13664 CCGAGCGCTGTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13717

QY 418 ACCCGAGCTGAGACCGCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 477
    |||||
Db 13718 ATCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13777

QY 478 CCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 537
    |||||
Db 13778 CGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13837

QY 538 CCGAGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 597
    |||||
Db 13838 CGAGAGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13897

QY 598 GTG 600
    |||||
Db 13898 GCG 13900
```

## RESULT 4

US-10-042-665A-3  
; Sequence 3, Application US/10042665A  
; Patent No. 6924106  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Toupet, Christine  
; APPLICANT: Engel, Nathalie  
; TITLE OF INVENTION: Kitamycin biosynthesis gene cluster  
; FILE REFERENCE: 4-21001/B/C1  
; CURRENT APPLICATION NUMBER: US/10/042,665A  
; PRIOR FILING DATE: 2002-01-09  
; PRIOR APPLICATION NUMBER: 09/242,744  
; PRIOR FILING DATE: 1999-03-24  
; PRIOR APPLICATION NUMBER: PCT/EP97/04495  
; PRIOR FILING DATE: 1997-08-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 53799  
; TYPE: DNA  
; ORGANISM: Amycolatopsis mediterranei  
US-10-042-665A-3

Query Match 16.3%; Score 98; DB 3; Length 53799;  
Best Local Similarity 53.1%; Pred. No. 7.7e-10;

Matches 256; Conservative 0; Mismatches 220; Indels 6; Gaps 2;

```
QY 125 CTTGGGCGCGGTGTCCCGGAGACCGGCGAGAGTGGCGGTACAGACTGCGCTGAGACCCG 184
    |||||
Db 13599 CCGGCGCGGAGCGCTCGACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 13658

QY 185 CGACTTCGCGGCTCTCTCTCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
    |||||
Db 13659 GAGAGCGAGAGCGCGCGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCAC 13718

QY 245 CCGGCGCTGAGAGAGAGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
    |||||
Db 13719 CACCGAGCGCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 13778

QY 305 GTGACCGGCGAGTGTGCGCGCGCGCGAGATCGGCGCGCACTGAGAGCGCTGAGAGAGAG 364
    |||||
Db 13779 GAGAGAGAGTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 13838

QY 365 CTTGTCACTGTGTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 421
    |||||
Db 13839 CTTGACCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13898

QY 422 CAGCTTGAACCTTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 481
    |||||
Db 13899 CACCGCGTGAACCTTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 13958

QY 482 GTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 538
    |||||
Db 13959 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14018

QY 539 CCGAGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 598
    |||||
Db 14019 CCGAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 14078

QY 599 TG 600
    |||||
Db 14079 CG 14080
```

## RESULT 5

US-07-642-734C-1  
; Sequence 1, Application US/07642734C  
; Patent No. 5824513  
; GENERAL INFORMATION:  
; APPLICANT: Katz, L  
; APPLICANT: Donadio, S  
; APPLICANT: McAlpine, J B  
; TITLE OF INVENTION: Recombinant DNA Method for Producing  
; TITLE OF INVENTION: Erythromycin Analogs  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Edward H. Gorman  
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
; CITY: Park Rd  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/642,734C  
; FILING DATE: 17-JAN-91  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Danckere, Andreas M  
; REGISTRATION NUMBER: 32652  
; REFERENCE/DOCKET NUMBER: 4952.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 708-937-9396

TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11219 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 2338  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 744..6659  
OTHER INFORMATION: /function="APPROXIMATE SPAN OF  
OTHER INFORMATION: /label=FUNCTION  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 744..11219  
OTHER INFORMATION: /function="gene="eryA"  
OTHER INFORMATION: /product="ORF1 encoding modules 1 & 2 for  
OTHER INFORMATION: 6-deoxyerythronolide B"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 744..1868  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: acyltransferase domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1998..2198  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: acyl carrier domain 1 of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 2250..3626  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: beta-ketolactam synthetase domain/module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3831..4811  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: acyltransferase 2 domain of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 5574..6125  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6369..6626  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: acyl carrier domain 2 of module 1"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6678..11219  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 6678..8066  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: beta-ketoacyl ACPSynthase of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 8262..9305  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: acyltransferase domain of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 9906..10454

OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: beta-ketoreductase domain of module 2"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 10707..10964  
OTHER INFORMATION: /function="approximate span of  
OTHER INFORMATION: acyl carrier domain of module 2"  
US-07-642-734C-1

Query Match 15.5%; Score 93.2; DB 2; Length 11219;  
Best Local Similarity 52.2%; Pred. No. 6.3e-09;  
Matches 311; Conservative 0; Mismatches 243; Indels 42; Gaps 3;

QY 1 CGCGCTGCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 60  
DB 9190 CGGACTGCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9249  
QY 61 AGGAGCGCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 120  
DB 9250 GGGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9309  
QY 121 TCCGCTGCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 180  
DB 9310 CGGAGCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9369  
QY 181 CGGAGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 240  
DB 9370 AGGAGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9429  
QY 241 TACCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 300  
DB 9397 TCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9456  
QY 301 TGTGCGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 360  
DB 9457 AGGAGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9510  
QY 361 CGGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 420  
DB 9511 TCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9570  
QY 421 CCAGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 480  
DB 9571 GTCGCGGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9630  
QY 481 TGTGCGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 540  
DB 9631 TGTGCGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9687  
QY 541 AGGCGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 596  
DB 9688 AGGCGAGGTAACCCGCGGGGTGAGGATCCATCGGCTACCGCAGTGGTG 9743

RESULT 6  
US-08-439-009A-1  
Sequence 1, Application US/08439009A  
Patent No. 6004787  
GENERAL INFORMATION:  
APPLICANT: Donadio, S  
APPLICANT: Katz, L  
APPLICANT: McAlpine, J B  
TITLE OF INVENTION: Method of Directing Biosynthesis of  
TITLE OF INVENTION: Specific Polyketides  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steven F. Weinstein  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
CITY: Park Rd  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952. US, D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..11219
OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
OTHER INFORMATION: MODULE 1"
OTHER INFORMATION: /label= FUNCTION
FEATURE:
NAME/KEY: CDS
LOCATION: 744..11219
OTHER INFORMATION: /function= "gene= "eryA""
OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
OTHER INFORMATION: 6-deoxyerythronolide B""
FEATURE:
NAME/KEY: misc_feature
LOCATION: 744..11868
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1998..2198
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2250..3626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3831..4811
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase 2 domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5574..6125
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6369..6626
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain 2 of module 1"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..11219
OTHER INFORMATION: /function= "approximate span of
```

```
OTHER INFORMATION: module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6678..8066
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8262..9305
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9906..10454
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 2"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10707..10964
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 2"
US-08-439-009A-1

Query Match 15.5%; Score 93.2; DB 3; Length 11219;
Best Local Similarity 52.2%; Pred. No. 6.3e-09;
Matches 311; Conservative 0; Mismatches 243; Indels 42; Gaps 3;

QY 1 CCGGCTGCGCGAGGCGTACACCCGGGGGTGAGTCTGACCTGACGCACTGGGATG 60
DB 9190 CGGCACTGCGCGACGCGCAACGCGCGGTGCGGTGACGCTGAGGAGACGGTGTGCGGCC 9249
QY 61 AGGACGCCCGGTGACCTGCGCGGTCTACCGGTTCCACGACAGACTTCTGCTCCCGG 120
DB 9250 GGGCGGACCTGTGACCTGCGCGGTATCTTCCAGGGAACGGGTTGCTGCTGCTGC 9309
QY 121 TCCCCCTGGGCGGAGTCCCGACACCGGACAGAGTGGCTTACAGCTGCGCTGGACAC 180
DB 9310 CGACGCCACACCCCTGCTGACAGAGTCAAGCGCTGCTTACCGGAGTCACTGACCG 9369
QY 181 CCGTGAACCTGCGGCGGCTCTCTGCGCGGACGGGCTCTGTGATGACCGGACGAG 240
DB 9370 AGGTGCGCGCTCCGAACCTGCG-----GCGG 9396
QY 241 TACCCCGGCGCTGACGCACTGTGTCGCGACGCGCTGGAACAGCGCGGCGACCTGTG 300
DB 9397 TGCGCGCGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9456
QY 301 TGTGTGACCGCGCAAGTGGCGCGCGCGGATCGCGCGCTGACCTGACCGCTGACGCGCA 360
DB 9457 AGGTGCGGTCG-----CGCTCGCGGAGCGCGCGCGGACCGGAGGTCAAGCGCGCG 9510
QY 361 CCGGCTGTCACTGTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
DB 9511 TCGGCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9570
QY 421 CCAAGCTGACACCTGCGGTTGCTCAAGCGCTGCGCGGACCGCGGATTCGACGTCCTCC 480
DB 9571 GTGCGGTGCAACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9630
QY 481 TGTGCTGTGTGACCAAGGAGCGCGCGCGCTGTGACCGCTGGAAGACGACGTGATCGGCGCC 540
DB 9631 TGTGACGGTCACTTGTGCGCGCGGTGACGCG---GGCAGTCCGCTGCGCGCGCGGAC 9687
QY 541 AGGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 596
DB 9688 AGGCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 9743

RESULT 7
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
```

```

1  APPLICANT: Sherman, D.H.
2  APPLICANT: Liu, H.
3  APPLICANT: Xue, Y.
4  APPLICANT: Zhao, L.
5  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
6  FILE REFERENCE: 600,438US1
7  CURRENT APPLICATION NUMBER: US/09/105,537A
8  CURRENT FILING DATE: 1998-06-26
9  NUMBER OF SEQ ID NOS: 43
10 SOFTWARE: FastSeq for Windows Version 3.0
11 SEQ ID NO 30
12     LENGTH: 13842
13     TYPE: DNA
14 ORGANISM: Streptomyces venezuelae
15 US-89-105-537-30.

```

Query Match	14.9%	Score 89.4	DB 3	Length 13842
Best Local Similarity	52.6%	Pred. No. 3.1e-08		
Matches 317, Conservative				
			0, Mismatches 271,	Indels 15, Gaps 5

QY	1	CCGCGCTCCGCGAGGCTGTAACAACCCGCGGGGGGTGAGAGTGCAGCTGCGGTGACCGCAGTGG---G	57
Db	5849	CCTCGCTCCGCGAGGCGTGGGGTCAACGCGGCTTCCTGGTGGCATGTGACTTCGCTCTCGCCCG	5906
QY	58	GTAAGGAGACGCCCGGCTGCACTGCGCGGTCTAACCGCTTCCAAAGACAAATTCTGAGCTTC	117
Db	5909	CCACGGGCTCTCCGCGCCCGGCTCTGGCCCACTTAACGCTTTCACAGGCCAGGCGCTACTGAGCTTCG	5968
QY	118	CGGTCCGCCCTGAGGCGCGGGTCCCGCAACAACGGAGCAAGATGGCGTTTACAGAGCTCCGCTGGC	177
Db	5969	AGAAACATCTCCCGCC---GCCCTGGCCCAACGGCGAGCACTGGCGCTTACCGCATGACTGTGA	6022
QY	178	ACCCTGCTGACCTTCGGGCGGTCTCTCCCTGAGCCGAGACGGAGTCTGTGTGTGACCCGAGCGCG	237
Db	6026	AGCGCTCTCCGCGCGCGCGAGGGGAGTCCGAGCGCACCG---CCTGTCCGGCGCGCTGGCTCG	6082
QY	238	CAGTACCCCGGCGCCCTGGACGAGACGTGTGTCCGCGAGACGCGCTGTGAACAAGCGCGGCGACCG	297
Db	6083	CCGTCAAGCCCGGAGGAGCACTTCGCGGCAAGGCGCGCGCGGTGTCTAACCGCGCTGGTTCGAGCG	6142
QY	298	TGCGTGTGTGACACCGCGCAGTCCGCGCGCGCGGATCCGAGGCGCGCACTTCAGACCGCGTTCGAGCG	357
Db	6143	CCGGGGCGAAGATGAGAGTGCCTGACGCGCGGGGCGGAGCAAGACAGCCTGTAAGGCTTCGCGCG	6202
QY	358	GCAACGCGCTGTTCACCTGTGTGTCTCTCTGCTCGGCGCTCGCGAGGCGCGAGTGTCTGACAG	417
Db	6203	---CCGGGCTCACCGCACTGACGAGACCGGTGAGCGGCTTCAACGGCGGTGTCTGCTCTCGCG	6255
QY	418	ACCCCAGCCTGAGACAACCTTCGCGTGGTTCGAGGCGCTCGGCGCAAGCCCGGATTCGAGCTCC	477
Db	6260	ACGGACCTTCGTAACCGCAAGTTCGCTGAGGTTCGAGGCGCTCGGCGAGCGCCGGAATTCAGGCGCGC	6319
QY	478	CCCTGTGAGCTGTGACACGAGGAGCGCGCGCGCGTGAACCG---TGGAGAACGAGCGTTCGATTC	534
Db	6320	CCCTGTGTGTCCGTCAACCCAGGCGCGCGGTCTCCGTTCGAGCGTCTTCAGACAACCCCGCGAGCC	6379
QY	535	CGGCGCCAGGCGCATGATCGTGGGAGCTCGGCGCGGATGTGTGGAGCGGTGAGATCCCGCGCCGAT	594
Db	6380	CCGACCGGAGCATGCTCTGAGGGGCGTCCGCGCGCGTGTGCGCCTTGTGAGACCCCGAAGCGCT	6439
QY	595	GGG GGG GGG	
Db	6440	GGG GGG GGG	

US-09-105-537-5  
 Sequence 5, Application US/09105537A  
 Patent NO. 6255202  
 GENERAL INFORMATION:  
 APPLICANT: Sherman, D. H.  
 APPLICANT: Liu, H.  
 APPLICANT: Xue, Y.

```

; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

```

Query Match	14.9%	Score 89.4; DB 3;	Length 36778;
Best Local Similarity	2.6%	Pred. No. 3e-08;	
Matches 317; Conservative	0;	Mismatches 271;	Indels 15; Gaps 5;

QY	1	CCGCGCTCCGCGAGGCTGTAACACCGGGGGGGTGAAGAGTTCACCTGGCCGTACCCCAAGT---G	57
Db	7590	CCTGCGTCGCGAGAGGCTGGGGGTCAACGGGGCTTCCCGTGGCATGTGACTTGGCTCTGCGCCG	7649
QY	58	GTGAGGGGACGCGCCGGTGCACCTGCGGGTTAACCCGTTCCACACGACAGAACATTCTTGCGCTCC	117
Db	7550	CCACGGGCGCTCCCGCGCCCGGTCTGGCCACTTACCGCTTCCAGGCCGAGCGCTTACCTGGCGTCG	7709
QY	118	CGGTCCCGCTTGGGCGGGGATCCCGCACACCGGCGACGAGTGGCGTTTACCAAGCTCGCGTGGGC	177
Db	7710	AGAACATCTCCGCGC---GCCCTGGCCACACCGGCGACGACTGGCGCTTACCGCATCATGTGGA	7766
QY	178	ACCCCGTGCACCTTGGGCGGGTCTCTCGTGCGGAGCGAGTCTTGAGTGAACCGAGCGG	237
Db	7767	AGCGCTCTCCGGCGCGCGAGGGGTCCGAGCGCACCGG---CTGTCCGGCGCGTGGCTG	7823
QY	238	CAGTACCCCGGCGCTGAGCGGAGCGTGTCCGCGACCGGCGCTGGAAACAAGCGCGGGCGACCG	297
Db	7824	CCGTACGACCGGAGAGACCACTCCGCGCAGGCGCGCGCGTGTCAACCGCGCTGTGTGACG	7883
QY	298	TGCGTGTGTGACACCGCGCAGTCCGCGCGCCGAGATCGGCGCCGCACTCGACGCGGTGACG	357
Db	7884	CCGGGGGAGAAAGTTCGAGGTCTGACCGGCGGGGCGGACGACGACCGTGGAGGCCCTTCCGCG	7943
QY	358	GCAACGCGCCTGTTCACCTGTGAGTCTCTGTGCTCGGCTTCGCGAGGAGCGGTGTGACG	417
Db	7944	---CCGGGCTCACCGGCACTGACGACCGGTGACGGCTTTCACCGGCGGTGGTCTCGCTCCG	8000
QY	418	ACCCCAACCTGAGAACACCTTCGCGGTGGTTCAGAGCGCTTCGGGCGACGCCGGGATTCGACGTCC	477
Db	8001	ACGGACTTCGACCGGCAAGTCCGCTGGGATTCGAGGCGCTTCGGCGACGCCGGAAATCAAGCGCG	8060
QY	478	CCCTGTGGCTGTGACACGAGGAGACGCGCGCGCCGCTGACCG---TCCGAGACGACGTGCATC	534
Db	8061	CCCTGTGTGTCGTGACCCGAGGGGCGGTCTTCGTGGGACGTTTCGACACCCCGGCCGACAC	8120
QY	535	CGGCGCCAGGCGCAATGATCGATGGGCTCGGCGGGGTGTGGCGTGGAGTCCCGCGCCGGT	594
Db	8121	CCGACCGGGCGCAATGCTCTGGGGGCGTCCGCGCGGTGTGCGCCCTTGAGACCCCGGAACGT	8180
QY	595	GGG 597	
Db	8181	GGG 8183	

RESULT 9  
 US-09-320-878-19  
 Sequence 19, Application US/09320878A  
 Patent No. 6117659  
 GENERAL INFORMATION:  
 APPLICANT: ASHLEY, Gary  
 APPLICANT: BETLACH, Melanie C.  
 APPLICANT: BETLACH, Mary C.  
 APPLICANT: MCDANIEL, Robert  
 TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASES



```

// FILE REFERENCE: 300622002120
// CURRENT APPLICATION NUMBER: US/09/320, 878A
// CURRENT FILING DATE: 1999-05-27
// EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
// EARLIER FILING DATE: 1998-08-28
// EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
// EARLIER FILING DATE: 1998-05-06
// EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
// EARLIER FILING DATE: 1997-04-30
// EARLIER APPLICATION NUMBER: 60/119, 139
// EARLIER FILING DATE: 1999-02-08
// EARLIER APPLICATION NUMBER: 60/100, 880
// EARLIER FILING DATE: 1998-09-22
// EARLIER APPLICATION NUMBER: 60/087, 080
// EARLIER FILING DATE: 1998-05-28
// NUMBER OF SEQ ID NOS: 34
// SOFTWARE: PatentIn Ver. 2.0
// SEQ ID NO 19
// LENGTH: 38506
// TYPE: DNA
// ORGANISM: Streptomyces venezuelae
// US-09-320-878-19

```

QY 595 GGG 597  
Db 6323 GGG 6325

RESULT 11  
US-09-657-440-19  
; Sequence 19, Application US/09657440  
; Patent No. 6509455  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30622002120  
; CURRENT APPLICATION NUMBER: US/09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 14.9%; Score 89.4; DB 3; Length 38506;  
Best Local Similarity 52.6%; Pred. No. 3e-08;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;  
QY 1 CCGCGCTCGCCGAGCGTACACCCGGGGGTGAGGTGACTGCGCTGACCGAGTG---G 57  
Db 5732 CCGCGCTCGCCGAGCGTACACCGGCTTCCGTCGATGACTGCTCTCTCCCG 5791  
QY 58 GTGAGGAGCGCCCGGTGACCTGCGGTCTACCCGTTCCAGACAGAACTTGTGCTCC 117  
Db 5792 CCAAGGCTCTCCCGCCCGGTGCTGCCACCTTCCAGCGCGGAGCGCTACTGCTCG 5851  
QY 118 CGGTCCCGCTCGCGCGGTGCTCCGACACCGCGGACGAGTGGCTTACCACTGCTGCGC 177  
Db 5852 AGAACACTCCCGC---GCCCTGGCCACCGCGGACGACCTGGCGCTACCGCATGACTGGA 5908  
QY 178 ACCCGCTGACCTCGGGCGGTCTCTCTGCGCGGAGCGGTCGTGATGACCGAGCGG 237  
Db 5909 AGGCGCTCCCGCGCGCGGAGGGGTCCGAGCGCACCGG---CTGTCCGGCGCTGCTCG 5965  
QY 238 CAGTACCCCGCGCTGAGACGAGTGTCCGCGACCGCTTGAAACAGCGCGGCGGACCG 297  
Db 5966 CCGTACCGCGGAGACACTCCGCGCAGGCGCGCGCGCTGCTCACCGCGCTGTGAGAG 6025  
QY 298 TGTGTGTGTGACCGCGGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357  
Db 6026 CCGGGGCGAAGGTGAGAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6085  
QY 358 GCACCGCGCTGTCATGTGTGTCTCTCTGCTGCGCGCTGCGCGGCGGCTGTGAGAG 417  
Db 6086 ---CCGGGCTCACCGGACTGAGACCGGTGAGCGGCTTACCGGCGGTGTGCTCTCTCG 6142  
QY 418 ACCCGAGCTGACACCTCGCGGTGTGCTCAGGCGCTGCGCGAGCGCGGATGAGTCC 477  
Db 6143 ACGAGTCTGTACCGGAGGTGCGCTGTGTGCAAGGCGCTCGGCGAGCGCGGATTAAGCGCG 6202  
QY 478 CCGTGTGTGTGACCGGAGAGCGCGCGCGCGCTGACCG---TCGAGACAGCTGTGATC 534  
Db 6203 CCGTGTGTGTGCTGTGACCGGAGGCGGTCTCCGTGAGACGTCCTGCGACCGCGCGCGCG 6262  
QY 535 CGGCCAGGCACTGTGTGTGTGTGCGGCGGCGGTGTGAGGCGGTGAGTCCCGCGCGCGGT 594

Db 6263 CCGACCGGCGCATGTGTGTGCGGCTGCGCGGCTGCTGCGCTTGAAGACCGCGAAGCT 6322  
QY 595 GGG 597  
Db 6323 GGG 6325

RESULT 12  
US-09-793-708-19  
; Sequence 19, Application US/09793708  
; Patent No. 6902913  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30622002121  
; CURRENT APPLICATION NUMBER: US/09/793,708  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: US 09/657,440  
; PRIOR FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: US 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: US 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: US 09/073,538  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/846,247  
; PRIOR FILING DATE: 1997-04-30  
; PRIOR APPLICATION NUMBER: US 60/134,990  
; PRIOR FILING DATE: 1999-05-20  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-793-708-19

Query Match 14.9%; Score 89.4; DB 3; Length 38506;  
Best Local Similarity 52.6%; Pred. No. 3e-08;  
Matches 317; Conservative 0; Mismatches 271; Indels 15; Gaps 5;  
QY 1 CCGCGCTCGCCGAGCGTACACCCGGGGGTGAGGTGACTGCGCTGACCGAGTG---G 57  
Db 5732 CCGCGCTCGCCGAGCGTACACCGGCTTCCGTCGATGACTGCTCTCTCCCG 5791  
QY 58 GTGAGGAGCGCCCGGTGACCTGCGGTCTACCCGTTCCAGACAGAACTTGTGCTCC 117  
Db 5792 CCAAGGCTCTCCCGCCCGGTGCTGCCACCTTCCAGCGCGGAGCGCTACTGCTCG 5851  
QY 118 CGGTCCCGCTCGCGCGGTGCTCCGACACCGCGGACGAGTGGCTTACCACTGCTGCGC 177  
Db 5852 AGAACACTCCCGC---GCCCTGGCCACCGCGGACGACCTGGCGCTACCGCATGACTGGA 5908  
QY 178 ACCCGCTGACCTCGGGCGGTCTCTCTGCGCGGAGCGGTCGTGATGACCGAGCGG 237  
Db 5909 AGGCGCTCCCGCGCGCGGAGGGGTCCGAGCGCACCGG---CTGTCCGGCGCTGCTCG 5965  
QY 238 CAGTACCCCGCGCTGAGACGAGTGTCCGCGACCGCTTGAAACAGCGCGGCGGACCG 297  
Db 5966 CCGTACCGCGGAGACACTCCGCGCAGGCGCGCGCGCTGCTCACCGCGCTGTGAGAG 6025  
QY 298 TGTGTGTGTGACCGCGGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357  
Db 6026 CCGGGGCGAAGGTGAGAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6085  
QY 358 GCACCGCGCTGTCATGTGTGTCTCTCTGCTGCGCGCTGCGCGGCGGCTGTGAGAG 417  
Db 6086 ---CCGGGCTCACCGGACTGAGACCGGTGAGCGGCTTACCGGCGGTGTGCTCTCTCG 6142





**This Page Blank (uspto)**





```
gene
928..2061
/gene="megDV1"
928..2061
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,4-Isomerase"
/db_xref="GI:10179847"
/translacion="WAVGDRRLGRBELQMARGLYWGANGDLYSMLSGRDDPTTW
YERLRAAGRGYPYASRAGTWWGDRHTAEVLADGFTHGPPDAARMQVVAHCPSAUSA
GPERFEYARTEDASVTVDADMLQOCARLVTLGSRFDLVNDFAREPVTALGTATA
LKGVDPRLRSVLTATVCLDAQVSPPOOLATBEQALFALDEIDAVTGRDAVAVGVV
AELAAVGNVAVLAVTELPRLAADLPETATRTVTSRTSPGVHLERTTASDRR
VGVVDVPTGSGVTVAANAARDPEFTDPDRFDVDRGDAAEILSSRGSRTDLDAV
ATLATAALRAAPVLPRLSRGSPVTRRRRSFVANGLSRCPVEL"
2072..3382
/gene="megDI"
2072..3382
/codon_start=1
/transl_table=11
/product="TDP-megosamine glycosyltransferase"
/protein_id="AAG13908.1"
/db_xref="GI:10179843"
/translacion="MRVVFSSMAVNSHLFGVLPLASAFQAAQHEVAVVAPALTDVY
GAGITAPVGDVDELVERMAHAGODIYEMPTLMDQSHHTMSMDLLCMQTETPTP
FFALMSPSLIDQVNEFCRWRPFIYMBPTTAPAPIMARTGTPIHAKMLGPPVATR
ARQSFRLLAHQEVREHREDPLAEFMDTLRRFGDPPLSEDEBELVQMTVDPIEPL
RIDGVTVGRVYVPYNGPSVVPAMLRRPERRVCLLTGSSRHHGIGQVSIEMLD
AIDIDASFEVTFDDQOLVGVSVAWVTRTGFPVNVALLPTCAVTHHGSGTSMILTA
AIIHGPOIILSDADTEVHAKOIDLQAGLSLPVAGMTAEHLRGAIEHVLDEPAVRLGA
ERHDDGKRTDPSPAQVNGICQDLAADAARQRORRTAEPLPR"
3462..4634
/gene="megY"
3462..4634
/gene="megY"
/codon_start=1
/transl_table=11
/product="mycarose O-acyltransferase"
/protein_id="AAG13909.1"
/db_xref="GI:10179844"
/translacion="WVTSINLDTTARPALNSLTGMRFFVAFLVFETHVLSRLINISYV
YADIDVEMQTTGRCVGSFFPIISGFUTWSAGSDVSMFRBRVCKLPFNHLYTAF
AAVUDPLVTGQAVSGBALIPULLIHAMPALBESFGINPYSKLACGAFYLCPEPLF
LFWISGIRBERLMAAAVFAIIVAVPVADLLPSSPPLIPGLSTSAIQWFLTFP
ATRSLEPTLIGIILARILITGRMINVGLPAVLLPVPFVASLFLGVVAISSMMILP
LVLIJASGATADLOOKRTFMERNRVVWMLGDVSPALVWVFLVIVGADLLGFSQTEDA
PLGLALFWIIFPLAVSLVLSMLVRFVELPVMRWARPASARRKATPEQTGPRR"
4651..5775
/gene="megDI1"
4651..5775
/gene="megDI1"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/protein_id="AAG13910.1"
/db_xref="GI:10179845"
/translacion="MTTYWSYLTLEYERERADILDVQKVPASGSLIIGQSENFETE
YARVYGHACVGVNDGTNAVKLATESVGVGDDERVVTSNTAPAPVLLAIDEIGARPV
ADVREEDYLDMDLVEAAVTPETKAIIVPHLYGQCVDMTALBELADRGLKVECAQ
AHGARDRDLAGTMSDAASFPYPTKVYAGDGAVVTNDETARALRLRYGMEV
VYVYTRTGRHSRLDEVGAELIRKULTLDAVAGRAVAGRYVDGLADLDSDHLEL
SVVTDGNEHVYVYVVRHPRDEIILKRLDGDIDSLNTSYPMVATMTGFALHVASG
SLPVTERLAGELFSLPMYPSLPHDIQDRVIEAVREVTIGL"
5822..6595
/gene="megDI11"
5822..6595
/gene="megDI11"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N-dimethyltransferase"
/gene
928..2061
/gene="megDV1"
928..2061
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"
/db_xref="GI:10179847"
/translacion="MPNSHETTSSTVAPYERADIVHDFYHGKGYRABADALVEVA
RKTPOALTLDAVCGSHLVELADSFREVCVDSAAMLATARNDPRLHQGM
RDSIDRRPDPVATCGSSGTGYLDBAEILDAVAVNAGHLAAGTIVVEPMPFETPRP
GMVGDVLTSGDKRISMSHITVPAGLDPDRTRASMTTHTYTGSPBAGIEHTBVHVMTL
FARAAAYEQAFQRRGLSCSYVDLFPGLFVVAAEGR"
6592..7197
/gene="megDI1"
6592..7197
/gene="megDI1"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
/translacion="MVVLGSGSGLFSGAVNTHALADLPYRVULVAREVYVSGAVADYE
THRDVLTPEALAEVADARAVERPFAAQIQIGTISQMRISBDDVVAERTNVGLVRLIYV
LBSRPAFVVPFSSNTQVRYAVRQIDBSBDHPGKGVDRDROHTEBQLKERTYAG
AIRATSLRLPVEFGVPAAGTADRGVSTIMIRALQOPLTMHMDGTVRRELLVYTA
ARAFTALDPAEDALAGRHFLGDRSGVSTLPRFALVRSVLRHGEDEPVPVSPPA
HMPDSLRSEVNDPARFTATYCWRAITYTMAEAVDRTVAALAPRAAAPSERS"
7220..8206
/gene="megDV"
7220..8206
/gene="megDV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
/translacion="MVVLGSGSGLFSGAVNTHALADLPYRVULVAREVYVSGAVADYE
THRDVLTPEALAEVADARAVERPFAAQIQIGTISQMRISBDDVVAERTNVGLVRLIYV
LBSRPAFVVPFSSNTQVRYAVRQIDBSBDHPGKGVDRDROHTEBQLKERTYAG
AIRATSLRLPVEFGVPAAGTADRGVSTIMIRALQOPLTMHMDGTVRRELLVYTA
ARAFTALDPAEDALAGRHFLGDRSGVSTLPRFALVRSVLRHGEDEPVPVSPPA
HMPDSLRSEVNDPARFTATYCWRAITYTMAEAVDRTVAALAPRAAAPSERS"
8226..8206
/gene="megDV"
8226..8206
/gene="megDV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
/translacion="MVVLGSGSGLFSGAVNTHALADLPYRVULVAREVYVSGAVADYE
THRDVLTPEALAEVADARAVERPFAAQIQIGTISQMRISBDDVVAERTNVGLVRLIYV
LBSRPAFVVPFSSNTQVRYAVRQIDBSBDHPGKGVDRDROHTEBQLKERTYAG
AIRATSLRLPVEFGVPAAGTADRGVSTIMIRALQOPLTMHMDGTVRRELLVYTA
ARAFTALDPAEDALAGRHFLGDRSGVSTLPRFALVRSVLRHGEDEPVPVSPPA
HMPDSLRSEVNDPARFTATYCWRAITYTMAEAVDRTVAALAPRAAAPSERS"
8226..9220
/gene="megDV11"
8226..9220
/gene="megDV11"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"
/protein_id="AAG13914.1"
/db_xref="GI:10179849"
/translacion="MGTTGAGSARVVRGSRALHTSRMLGTVNFSGRYTDDALRLMD
HALERGVNCTIDATIVGMRLYKGTETELVGVWFOGGGRRETYLATVYGSMBERVN
DGGISAHNIYVAAAGNSRLRGVNDIIDYOTHIIDRAAPMBWQAAEHLVSGSGVGVY
GSSNLACMHIYAAQESARRLNLGHIHQCLYNLAHNPBELDVIPAAQAYGVGPYAWS
PLHGSLGVLAEKLAGTAVKSAQRAQVLLPAVRPLVEAYEDYCRRLGADPAEVLGA
WVLSRPGILGAVIQPRTPEQIDSLRAAEYTLGSEBDELEBAIFPAPAVDGPVP"
9226..10479
/gene="megBV"
9226..10479
/gene="megBV"
/codon_start=1
/transl_table=11
/product="TDP-mycarose glycosyltransferase"
/protein_id="AAG13915.1"
/db_xref="GI:10179850"
/translacion="MRVULISFARHTFQGLVPLAMALHTAGHYRVASQBELTDVVV
GAGLTSVSGDHLFDISPEAAQVHRVYTTDLDFARQGLRKEWELFNGIEBARSF
VFPPVNNDSFVDEIVEFAMDMRPDLVIVMEPTFAGAVALAKAGAAHALLMGSDLTGY
FRSNDLROGRRPADDEPDPGLGMLTEVAGRFGLDYSDLLAVQMSVDOLPESEFLAT
GLSEVHRTTLPSYNSVVPOMLRTSGDGRVCFTGGYSALGITSNPOEFLRTLIAR
PDGEIVTTRSGDASVPDNRILVDVPMNILLPGCAVITHGGAGSNA TALHGVPO
ISVAHEMDCLARGRTAEIAGGVFLRPEVDADTLTMQALATVVEDRSIAENAKLROE
ALAAPTEAVVPEVLAALHQHADR"
10483..11424
/gene="megBI1"
10483..11424
/gene="megBI1"
/codon_start=1
/transl_table=11
```



/transl\_table=11  
/protein\_id="CAC37809.1"  
/db\_xref="GI:13938930"  
/translation="MTTVYVSYLLEVERERADITLDAVOKVFAAGSLILIGSOVENFET  
YARVGHIAHCVGDNKNTNARKLALIESGVRDREVVYSTAPATYLAIDIEGARPP  
VADREDEYLDMDTLEAVALPRTKALVPHLYGQCVMTALRELDAGRGKLVEDCAQ  
AHGARBRGLAGTMSDAAFSPYTKVLAAGGAAVYVNDDEFAARLRRLRIGMBE  
VYVYTRTPGHSRLDEVOAELRLKRLDLAAYAGRAVAKORVYDGLADIQDSHGL  
PVVDGNEHYFVYVVRPRDEIKLRDGYDLSINISYPMVPMHTGTGAHGVASG  
SLPVTERRLAGRIFSLPMYPSLPHLDQDRVLEAREVITGL"  
5822..6595  
/note="unnamed protein product; megDIII,  
daunosaminyl-N,N-dimethyltransferase (eryCVI homolog) ; SEQ  
ID NO: 7= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37810.1"  
/db\_xref="GI:13938931"  
/translation="MPNSHTSSTSDVAPYERADIVADFYHGRKGGRABADLVEVA  
RKHTPOATLIDVACGTGSHVELADFREVGVDSLAAALATAPARDPGRPELQGM  
RDSLRDPVYVCMPSSTGYLNDREALDRAVANLACHILAPGTLVPEPMPETFP  
GMVAGADLVTSQDRISRMSTHYVPAGLDRTRRSRTTHYTTGSPAGLIEHTVAVMTL  
FARAYEAFQFQAGLSCSYVGHDLFSFGELFVGAABEGR"  
6592..7197  
/note="unnamed protein product; megDIV,  
TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU  
homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase; SEQ ID  
NO: 8= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37811.1"  
/db\_xref="GI:13938932"  
/translation="MRYVEGLTBEVFTPTPTPADERGVRGTAYOEDVFAALGRPLF  
PVAQVSTRRGVYRGVHTFTTNGSMKATYTCARGAAMDPAVDIRGSPFPGAAEY  
ELSAENWGLYLPVGMGHLFVSLDDDTLVLMGAGVYDPDEKRAVHPUDPELALPFA  
DLIVMSERDRVAFTLRARDOGLIPDYAACRAAHRVYR"  
7220..8206  
/note="unnamed protein product; megDV, TDP-hexose  
4-ketoreductase (eryBIV, dnmV homolog),  
TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37812.1"  
/db\_xref="GI:13938933"  
/translation="MVVVGASGFLGSAVTHALADLPYRVYLVARREVVVPSGAVADYE  
THRVLTPEGALAEVADARAVPFAAQIRGTSGMRISSEDVVAERTNVGLRLIAY  
LSRSPHAPVVVPGSNTQVGRVTVAGVIDGSEBQHPBEVYDRKHGTQHOLKEXATAG  
AIRATSLALPVFPGVPAAGTADRGVATMIRALATGQPLTMDHGTGVRRELIVYTA  
ARAFTVTLDDHALAGRHFLITGSRSMWLGIVFOAVSRSVARHGEEDVPVAVPPPA  
HMDPSDLRSVAYDPARFTAVYGMRAITVMAAANDRTVAALAPRAAABSEBS"  
complement(8228..9220)  
/note="unnamed protein product; megBII-1(megDVII),  
TDP-4-keto-1,6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37813.1"  
/db\_xref="GI:13938934"  
/translation="MGTTGAGSARVVRGSRALHTSRLMIGTVNFSGRVTDDALRLMD  
HALBRGVACIDTADIVGWRLYKGHTBELVGMFAAGGGRRETVATLVGSEMSRVN  
DGLSARHIVAAEENSRLGVDHITDIOTHHIDRAPELDDVAEHLVSGVGVY  
GSSNLAGHIIAAOBSARRNILKMTSHOCLYNLAVRPEDVDLPAAQAYGVPAWS  
PLHGLLSGVLEKLAAGTAVKSAQGRAOVLPAVPLVEAIEDYCRKAGAPAEVGLA  
WLSRPGILGAVTGRTEPEQLDSALRAELTLGEBELRELBAIFPAPVNDPVP"  
complement(9226..10479)  
/note="unnamed protein product; megBV, mycarosyl  
transferase, mycarose glycosyltransferase; SEQ ID NO: 11=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37814.1"  
/db\_xref="GI:13938935"

/translation="NEVLITSPAHRTHEGGLYPLANALHTAGHDVAVASOPELTDVV  
GAGLTSVPDSHRLDFDISPEAAQVHARTTDDFARRKBPESLMEFLIGESTATSE  
VFVGVNNDSEFVDELVEFAMDWRPDLTMEPEFAGVAAKCAAAARALHIGSDLTGY  
FRSRGODLGGORPADRPDLGMLTEVAGRGFLDSEDLVAQMSVDDLPESFRLT  
GLESVHTRTLPGNNGSVVPOMLRTSGVRCVGTGYSALGITSNPOEFLRTATLAL  
PGEIYVTSGLDPAISVPNNVRLVDPNNILIPGCAVYHHGAGSMATLHGVDO  
ISVAHMDVLEQRTAEIGAGYFLRPDEVDATLMQALATVVEDSHENAKLQCE  
ALAAPPAEVAPELALAHQHRDR"  
complement(10483..11424)  
/note="unnamed protein product; megBIV, TDP-hexose  
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 12= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37815.1"  
/db\_xref="GI:13938936"  
/translation="MTRVYTLIGSGFVGSALLREFTHPLRLAVARTGSRDOPGS  
AGIEHRLVILDEGRVAQVYADTDVYVHLYAVAAAGSTYMSATVPEAREVNGIMD  
LVALARPQPAVLLFASTQANPAAPBRYAQHKIEARITLRQATDGVVDGVLIR  
LPATVGHSGSPGQGTGVVYAMTIRALAGPITMHGSRVRLMLHVEDVATFTAL  
HNHEALVGDVWTSADAEAPLGEIFETVAVSVARQGNPAVPVAVSPPPNABANDR  
SDPDFSTREFRTLGMHPRVPLAGIDRTVALISTXE"  
12181..22821  
/note="unnamed protein product; megAI; SEQ ID NO: 13=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC37816.1"  
/db\_xref="GI:13938937"  
/translation="WVDVPLDLGTRTPHPPELPFPMPICGHNEPELRAROLHAYLE  
GISBDDVAVAGALARETRAQDPSHRAVVVAASVTELTALALAQGRHPSVVRGYA  
RPTAPVFLVPGQAQMPGMATRLALIESPFAAAMCEAPBEVDMLSTLEVDSPE  
HLRREVEVOPALFAVOTSLAALMRSFGVRDPAVIGSILGILAAVCGADEVAAVA  
AALMSREMPVLVARGDAVALAPAEALAAVEEMDDVAGVNGPSVLLTGAPEBI  
ARVAYELAAQVPAQVYVNVNVAASQVDAVAGMSALTWMPAGSDVYVGLTGG  
RLUTRELGAGHPRFSTFLYPRFBATRAVLELPQFTIESPHVPLPSAVAYETDGB  
PSTEFDAADPDYLRARALREIYAEALALAGREVDATATREGLDSVLAVOLTRIA  
TATGRDLHIMLYDHPFALTEALRGPDEBERGETHPTBAEDVAVAVAMAC  
RLPGVTSPEFWEELABGRDAVGLEPTDGMPLDSLPHDPPTSGTAHQRAAGFLTG  
ATSPDAPFRLSPRELAIVRPOORTITELSEVLEERAGIPTSLRTSGVYVGLIQ  
EYGRPLABGSEGVETGMTTGTTSVAGRAYITLGLGPAISVDPAASSSLVAVHAC  
QSLRGERSTHALLAGVTVMPPTKMLVDPSMNSLAPGRKSAFSAALDGMAGBAGH  
LLERLSDARRHGHPLVAVIRGTAVNSDNGLSAPNGRAQVRVRLQALAESGLTGH

Query Match 100.0%; Score 600; DB 6; Length 47981;  
Best Local Similarity 100.0%; Pred. No. 36477;  
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCGCGCTCGCGGAGCGTACACCCGGGAGGTGAGTCTGACTGCGCGTACCGCACTGGGCTG	60
DB	36307	CCGCGCTCGCGGAGCGTACACCCGGGAGGTGAGTCTGACTGCGCGTACCGCACTGGGCTG	36306
QY	61	AGGGAACGCGCGGTGACCTGCGGCTTACCCGTTCCAAACGACAGAACTTCTGGCTCCCG	120
DB	36367	AGGGAACGCGCGGTGACCTGCGGCTTACCCGTTCCAAACGACAGAACTTCTGGCTCCCG	36426
QY	121	TCCCCCTGGGCGGGTCCCCGACACCGGCGACAGTGGCGCTTACCAAGTGGCTGGACAC	180
DB	36427	TCCCCCTGGGCGGGTCCCCGACACCGGCGACAGTGGCGCTTACCAAGTGGCTGGACAC	36486
QY	181	CCGTGACCTGGGCGGGTCTCCCTGAGCGGACCGGCTCTGATGATGACCGGACGGCGAG	240
DB	36487	CCGTGACCTGGGCGGGTCTCCCTGAGCGGACCGGCTCTGATGATGACCGGACGGCGAG	36546
QY	241	TACCCCGGCGCTGAGCGGACGTGTGTCGCGACCGGCTTGAACACGCGCGGCGGACCGT	300
DB	36547	TACCCCGGCGCTGAGCGGACGTGTGTCGCGACCGGCTTGAACACGCGCGGCGGACCGT	36606
QY	301	TGTTGTGACCGCGGACGTGCGGCGCGGACGTGCGGCTGACCTGACCGGCTGACCGGCA	360
DB	36607	TGTTGTGACCGCGGACGTGCGGCGCGGACGTGCGGCTGACCTGACCGGCTGACCGGCA	36666

[illegible]

```
/function="approximate span of beta-ketoacyl ACP synthase
domain of module 4'"
misc_feature
6054..7026
/gene="eryA"
/feature="putative"
/function="approximate span of acyltransferase domain of
module 4'"
7165..9216
/gene="eryA"
/feature="putative"
/function="approximate span of dehydratase and
enoylreductase domains of module 4'"
9433..9984
/gene="eryA"
/feature="putative"
/function="approximate span of beta-ketoreductase domain
of module 4'"
10225..10483
/gene="eryA"
/feature="putative"
/function="approximate span of acyl carrier domain of
module 4'"
10723..20235
/gene="eryA"
/feature="ORF3 encoding modules 5 & 6 for
6-deoxyerythronolide B formation; putative"
/codon_start=1
/transl_table=1
/protein_id="AAA26495.1"
/db_xref="GI:152695"
/translation="MSGDNQTEBKLRLRYLKRVTYELDSVTARLREVEHRAEPIAV
GMACRFPGVDSPSPFMRVSGGSDALTAEPADGMPDDARGLAAAGDDAF
FGISPRALAMDPOQRIMLISMEALERACHDPSLGSNGVTGVTGVDYDPR
APDEVLTVGTGASVVASGRVATCLBEGMPTVDYACSGLTGALNAMESLRBEC
GLALAGVTWSSGAPTEFRSQGLADCKPFAAGFGIABGAVLVQRISA
ARRGRVLAIVRSGAVNODASNGLTAPSGPAQORYIRALBNAGADVIVYEAH
GTRGLDPLIVTHALLSTYOGARDPDDPLMTGSVSKYIGTQALAAAGVAMVLAIR
HGMPTLHFDEBSPQIEMDLSAVSVSOARSPAPAGVSSGIGTGTTLAT
BBAEADPEPPAPDSBPVLYISGRDQARQAQGRADHARPRNSLIDGTTLAT
RRSABERRAAVVGDRDALGLRAVAGRLADRTATQARTRRGVAVFPQGMQMG
MARDLRESOVFADISIDCERALA PHVWSLTDLTSGARPDRVVOPALFAVWSL
AALMRSHGVPAAVVGSQGLIAAHVAGLTLEDAKLVAVRSVARIQGGQWMS
FGIETBOAERIIGRFAGALSIASVNGPSRVVAGESPDEBELMECAEGITARIIV
DYASHSPVSLIRELLTELGISPVASDALVSTTGOPIDTATQATVYANLRQ
VREDDATROLAENAFDAFVSPHPTVTGIEATLDSALPDADGACVGTGTRBGGT
ADFTALGEATAGCVBVDWSPFADAPFVLPVTFQQRWTMLPIPGCARDEDDM
RYQVWRBAEWESASLAGRVLLVTGPVPSLSDAIRSGEAGATVLTCDVESRSTI
GTALAEADDTALSTVSLISRDEAVDPISDALVQALGAAGVAPLWVLTENAQV
ADGELVDPQAMVGLGRVVGIEQGRMGGLVDLVDADASIRSLAALADPRGECV
AIRADGIKVARIYAPARAATRWSPRGTVLVGTGIGIHAHVARMRLARSGAEHLVL
GRGADAPGASLEAEHLALGTCTTAAACVADARLRAVLAARBAGRVTSAYMHA
GVSTSTPDLDTLEAFTEIADVKVGTVNLDELCPDDDAFLPSSNKGWSPGLASY
AAANAFLDGFARRRRSGAPVTSIANGMAQONNAGBGGYILSGIRAPDPRAVE
ELHITLDHGQTSVSVVMDRRRFEVLFTHAHRPLFBEIAGARLEAOSGEPALOR
LAASTLRREHRLHLIRAEVAALVGHGDDAIDRDAFRFLGDSMTVAIDLRLAA
VTYGRBAATVFHDPITTRLADHYLLELVGAEEQAPALVREVPKXADDPILVGA
CRPFGVHNPCELMFEYVGRDATTBPTDGMULDLPDDPRRHTSYRRHAFID
GAADFAPFQISPREALAMPQOROVLETTWELFENAGIDPHSLKSDVTFGLAAT
QGYGODAAVPEBDSGYLLTGNSSAVSGRVAVYVIGBGPATVITACSSSLVALSAC
GLRLDQGLAVAGGVMAQPEVTEFRSQGLAVDGRCAFPSEADGFGFAGVAV
VTLRLSDRCGLAGROVAGVAGSAINDQAGNGLAAPGVQCAFQVIRAKMARBITGA
DVAVYEAHGTCTRLGDPYRASALATYKSGSSGSPVLTGSVKNADIGVADRAVAG
IKVVLGNLGLVPPMLCNGERSPLIEMSSGVLEAASVPPRACIDGRRAQVAFV
SGTNAHVIAEPPEPEPLPEBPGVGVLAANSVPLLSARTEHLAAQARLLEAVDV
SVPLTALASALATGARHLPRRALILAGHEBQRLQRLVAGVAGVAPGATTTGASLER
VFPFGQAGQWEGARGLISVPFAESIAECDAIVSEVAGSASEVLEQORDASLER
VDVIOPTLFSVMVSLARIMGACVSPSIVTGHSGEITAAVAVGLSLEDGRVLAIR
AKARIALAGKGMVSLAIPGRARALIPWDRISVAVAVNSPVSVSGDEBALAELV
AKCBDEGRAKTLVPDVASHSRHVEILSETLADIDGISARAAIPLVSTLGERRDG
ADMGRVYDMLRSQVRPDEAVSAADVAGHATFEMSHPVPLTAVOEILADAVAGIS
LHRDTAEHLIAELARAHVAVAAVDMRVFPALPPLPNPFEFORWMLAEVSDOL
ADSRVDMRPPLATTPVLEBGFVHGAPSLTSAVAKAGRVVVPVVSADREASAL
REVQEVAGVTSVHTGATHLAHLQSGEAGVAPIMVLTBRAVALGSESEVDEQAM
```

```
Query Match 47.8%; Score 286.6; DB 1; Length 20235;
Best Local Similarity 69.7%; Pred. No. 2.2e-18;
Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;

Qy 1 CCGCGCTCCGCGAGCGGTACCCCGGAGGTGAGGTGACCTGAGCTGACCTGAGTGGTG 60
| |||||
Db 13304 CCGCGCTCCGCGAGCGGTACCCCGGAGGTGAGGTGACCTGAGCTGACCTGAGTGGTG 13363
| |||||
Qy 61 AGGAGCGCCCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACGAACTTGAGTCCCGG 120
| |||||
Db 13364 ACCGCGGCGCGGTGACCTGCGCGGTCTACCCGTTCAAGACAGACGAACTTGAGTCCCGG 13423
| |||||
Qy 121 TCCCGCTGAGC---CGGATCCCGACACCGGCGAGGTGCGGTACCACTGCGCTGAGC 177
| |||||
Db 13424 TCCCGACCGCGCGGCGCGGACGAGGACGAGGACGACGACGACGACGACGACGACGACG 13483
| |||||
Qy 178 ACCCGCTGACCTGCGCGGTCTCTCTGCGCGGACGAGGTCTGAGTGTGACCGAGCGG 237
| |||||
Db 13484 GGAAGCGGAGTGGAGAGCGCTTCGTCGCGGACGCGGTGCTGCTGAGTACCGGACCGG 13543
| |||||
Qy 238 CAGTACCCCGCGCTGAGCGAGGTGCTGCGGACGCGGCTGGAACAGCGCGGCGAGCG 297
| |||||
Db 13544 GCGTGCCTGCGAGTGTGCGAGCCCATCCGAACTGACCTGAGACAGCGGTGCGACGG 13603
| |||||
Qy 298 TCGGTGTGTGACACGCGGCACTGCGCGGCGCGGATCGCGGCACTTGAACGCGCTGACG 357
| |||||
Db 13604 TCCGACCTGCGAGCTGAGTATCCCGTTTGAACATTCGACACGCGACGAGCGCGGACGA 13663
| |||||
Qy 358 GCAACCGCGCTGCTCACTGTGCTCTCTGCTCGCGCTGCGCGAGCGGTGCTGACG 417
| |||||
Db 13664 CCGACGCTGCTGCTCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13717
| |||||
Qy 418 ACCCGACCTGAGACCCCTGCGGTGCTGAGGCGCTGCGGAGCGCGGAGCGGAGTGAAGTCC 477
| |||||
Db 13718 ATCGGTGCTGAGACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13777
| |||||
Qy 478 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537
| |||||
Db 13778 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13837
| |||||
Qy 538 CCGAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 597
| |||||
Db 13838 CGAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13897
| |||||
Qy 598 GTG 600
| |||||
Db 13898 GCG 13900
| |||||

RESULT 4
AR049368 AR049368 20235 bp DNA linear PAT 29-SEP-1999
LOCUS
DEFINITION Sequence 3 from patent US 5824513.
ACCESSION AR049368
VERSION AR049368.1 GI:6005407
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz, L., Donadio, S. and McAlpine, J. B.
TITLES Recombinant DNA method for producing erythromycin analogs
JOURNAL Patent: US 5824513-A 3 20-Oct-1998;
FEATURES
source
Location/Qualifiers
1..20235
/mol_type="unknown"
/organism="unknown"
/feature="unclassified"
```

```
Query Match 47.8%; Score 286.6; DB 6; Length 20235;
Best Local Similarity 69.7%; Pred. No. 2.2e-18;
Matches 420; Conservative 0; Mismatches 174; Indels 9; Gaps 2;
```



QY	1	CCGCGCTCCGCGAAGCGCTAACACCCGGGGGGGTGAGAGCTGACCTGGCGCTAACCGCAGTGGGGT	60
Db	13304	CCGGGCTTCGGCGAGGCGCTTACCGCGAGGGCGTGGAGGCTGACTGAGACCCCGGCTTGCGCG	13366
QY	61	AGGAGCGCCCGGTGACCTGCGCGGTCTAACCGCTTCCACGACGAACTTCTGGCTCCGG	120
Db	13364	ACGCGCGGCGCGGTGACCTGCGCGGTCTAACCGCTTCCACGCGGACGCGGTACTGGCTGGCCA	13422
QY	121	TTCCCTCTGGGCG--CGGGTCCCCGACACCGCGACGAGTGGCGCTTACCACTGGCTGGC	177
Db	13424	TTCCCAACCGGCGGGCGCGCACGAGGACGAGGACGACGACTGGCGCTTACCAAGTGGTATGGC	13483
QY	178	ACCCGCTGACCTGAGGCGGGTCCCTCCCTGGCGCGGACGAGGCTGAGTGGTGAACCGGAGCGG	237
Db	13484	GCGAAGCGGAGTGGAGAGAGCGCTTCCGTGGCGCGGACGCGTGTGTGTGATCCGACCGG	13543
QY	238	CAGTACCCCGCGCGCTGACGCGACGTGGTCCGCGACGCGCTGGAAACAGCGCGGGCGACCG	297
Db	13544	GCGTGGCGGTCCGAACTTGTCCGACCGCATCCGAATGGATCTGGAGCAGAGCGGTGCCAGG	13602
QY	298	TGCTGTTGTGACCGCGCAGTCGCGCGCGCGGATTCGAGCGCGGACCTGACGCGGTGACG	357
Db	13604	TCCTGACCTGGGAGTGAATCCCGTTGACACATCGGCAACCGCATGGAGAGCGCGGACGA	13663
QY	358	GCAACCGCGCTGTCCACTGTGGTCTCTCTGCTCCGCGCTCGCGGAGGCGGTGTCTGACG	417
Db	13664	CCGACGCTGTGTCACTGTGGTGTCCGTGCTGTCCGCG-----ACGGGAGGCGGTGTG	13717
QY	418	ACCCGACCTGGACACCTCTGCGCTTGGTTCGAGGCGCTCGCGGAGCCGAGATGACGTCC	477
Db	13718	ATCCGTGGCTGTGACGCGCTCGCCTGTTCAGAGCCTTCGAGCGGCGGGGTGAAGCAC	13777
QY	478	CCCTGTGCTGTGACCGAGGAGCGCGCGCGGTGACCGTTCGAGACGACCTTCGATCCG	537
Db	13778	CGCTGTGGGTGTGACCGGACGCGGTGAGGTTGGCGGACCGGAACTGTGATCCG	13833
QY	538	CCGAGGCGCATGTGTGGGTCTGGCCCGGCGGTGGTGGCGGTGAGTCCCGCGCGGTGG	597
Db	13838	CGAGGCGCATGTGTGGGCGGTCTCGGCGCGCGGTGTGGCATTCGAGCAGCGGGCGCTGGG	13897
QY	598	GTG 600	
Db	13898	GCG 13900	
RESULT 5			
LOCUS	AR095529	20235 bp	DNA 1linear PAT 08-SEP-2000
DEFINITION	Sequence 3 from patent US 6004787.		
ACCESSION	AR095529		
VERSION	AR095529.1	GI:10023473	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 20235)		
AUTHORS	Katz,J., Donadio,S. and McAlpine,J.B.		
TITLE	Method of directing biosynthesis of specific polyketides		
JOURNAL	Patent: US 6004787-A 3 21-DEC-1999;		
FEATURES	Location/Qualifiers		
source	1..20235		

	ORIGIN	
Query Match	47.8%	DB 6; Length 20235;
Best Local Similarity	69.7%;	Pred. No.2,2e-18;
Matches 420; Conservative	0; Mismatches 174;	Indels 9; Gaps 2;
OY	1 CCGGCTCCGCAGGCGTACACCCGGGGGTGAAGTCGACTGCCATCCGAGTGCGTG 60	
Ddb	13304 CCGGGCTCCGCAGGCGGTAGGCCCAAGGCGTGAAGGTTCATCTGGACCCCCGCTTCGCGC 13563	

[illegible]

RESULT	6
LOCUS	AY661566
DEFINITION	AY661566 32299 bp DNA linear BCT-05-NOV-2004 Saccharopolyspora erythroaea polyketide synthase gene cluster, complete sequence.
ACCESSION	AY661566
VERSION	AY661566.1 GI:55419414
KEYWORDS	.
SOURCE	Saccharopolyspora erythroaea
ORGANISM	Saccharopolyspora erythroaea Bacteria; Actinobacteridae; Actinomycetales; Pseudonocardiales; Pseudonocardiaceae; Saccharopolyspora. 1 (bases 1 to 32299) Kodumal,S.J., Patel,K.G., Reid,R., Menzella,H.G., Welch,M. and Santi,D.V. Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster Proc. Natl. Acad. Sci. U.S.A. 101 (44), 15573-15578 (2004)
JOURNAL REFERENCE	2 (bases 1 to 32299)
AUTHORS	Hu,Z. and Reid,R.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUN-2004) Korean Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES	Location/Qualifiers
source	1..32299 /organism="Saccharopolyspora erythroaea" /mol_type="genomic DNA" /db_xref="taxon:1836" 1..10638 /gene="eryAI" 1..10638
gene	
CDS	



/gene="eryAI"  
/note="DBS1"  
/codon\_start=1  
/transl\_table=11  
/product="EryAI"  
/protein\_id="AAV51820.1"  
/db\_xref="GI:55419415"  
/translation="MADLSKLSDSRTAQPRIVRPMPLSGCNESALRARARQLAHLD  
RFPDAGVEGVAALAHDEQADAGHRAVVAASSTELDDGLAAVADTPHASPVRGVA  
RSPAPVVFVPGGAQOMAGMAEGLHRSVRPAAMDCARAFBYVDWTMLAVIDSRE  
QSRREVYQALRAVQTSIALNRSPGVTDVAVGHSITGELAAHYCGAAGDAANA  
AALSRKMTPLVNGDMAVALSADIEPRILAKMDVVLAVNGPERSVLTSSPEV  
ARRVSELSAGVRAQVINVMAASACQVDIAQESRSLAMFAGSGSEVFYTSITGG  
AVTRELVAHYMERSFLPYRFDAIRISALEVGTQFVEASPHFVLAALQOQLDAGE  
SSAAVPTLQRGQGMRFLLAAQAFITGVAVDMTAAVDVGEPSGLSALQPTAAEER  
DEPASGRTVNIALPHPTPRALAEALAACTEVAQRETRAKTEALRGEVAVVAMA  
KVSAGRTREYNIALYHPPTPRALAEALAACTEVAQRETRAKTEALRGEVAVVAMA  
CRLPGVSTPEEFWELISEGDVAAGLPTDRGMDDLSLFHPDPTRGSTAQRGCGFLP  
EATPFDPAFGMSPREALAVDPOQRMLIELSWEVLERAGIPPTSLQASPTVGGLTP  
QEVGPRLAEGGEGVGLMTGTTTSSVAGSLAYTLGLEPAILSVDTACSSSLVNHIA  
COSLRGESSILAMAGVTPMTPGMVDFSPMNSLAPDGRCAKASAGANGMAEGAG  
MLLERLSAPRNGHPVLAVIRGTAVNSDGSANGLSAPNGRAQVRVYIQAALSSGCGP  
ADIDVAHAGTGTGDPTEARRALPEAYGDRREPPLHGSYKSNLGTQAAGVAVTI  
KMYLAMPAGTLPTLHASERSKETIDSSGAILSLDEPEMPAGRPRAAGVSFGISG  
TNAAHIIIEAPQVVEGSEVAGDVAVPWLVSASBGLRQAQALAAHLRHPQODER  
DIAYSLATGRALPRAAFAPVDESALRVLDGLATGNADGAAGTSRAQORAVFVPP  
GQGMWAGMADVLLDTSPPVPAALRECDALPEHLDPEVITPLTEAAREEDAAIST  
BRVUNVOPMWAVWVSLASMMRAHGVPRPAIVGHSIGCEIAAACVAGALSLDDARVVA  
LRSRVITATRONKMASTIAPAGVTRPAGVITAVNGPERSVVAAGSDDELRLVA  
SCTTECIARRKADVASHSHVETITDALLAEIGDFHPLPGVFPFSTVGTGWTGP  
DELDAGVYMLRRTVRPADAVRALAOCYRTFLEVSAGHPLTAAIEIIGDGSADLS  
AIIHLRIGDGSILADFGALSRAFAVAVMESVHLGTGARVPLPTYPGQREVMLE  
PKPARSRETVDEVSALRYRIEMRPAGVEPARLDTGMVAKVAGTADENTSAREALE  
ESAPARRELVYVDAQCRDDELAERLSAGVAGVLTSLAVDEARENEALALASLT  
LSIQAVNSAELCCPLMTYTESAVATPEFFERVRAAGALMGVCRVTALENPAWGEI  
LVDPAGSVAELARHLAAVSGSGAGEDLALAKDSVYGRWRVRAAPATDEWKTGTV  
LVYGTGTVGGQIARMLARRGAPHLVLSRSGPDADAGEIVALEALGATITWTAACD  
VTDRSEVRELLGGIGDVPVLSAVHAAATLDDGVDITLGTGERISARAKVAGRNHL  
ELTRELDLTAFLVFSFSPASAFAGLGGVAPGNAYLDDLOQRSRGLPTATVAGRNH  
AGSGMAGPVADRPBRHGVTEMPEACRALQNALDRAENCPVITDVRMPRTLLAYRA  
ORPTRLDEIDARRAAPQAAEPFRVQALASLPBEREKALFEIVRSHAAVTLHASA  
ERVAPDAFAELGVDSLSELELRNLCAATGVRLPTTVFDPHPVTRLAAHLAEIGS  
ATGAEOAPATATVADBPITALVQNAICPLGCVDSPEERIMELITSGRDSAEVPPDRM  
VPDELMASDAGTRARAGNFWAGAGDFDAFFGSPREALAMDPOQOALETTEALE  
SAGIPPTERTSGDTPGVGMGHOGYATGRPPEPDGCGYLLTGNTASVASGRIYVIG  
LEGPALVYDTRACSSLYALHTRACSLDGDGGLAVAGSVMAPEVTFEERSGALS  
PDGCRKPPSDADPFGLGESAPVULQRLSDAREGRVLSVAGSAVNDQASGSGS  
APSGVAQORVIRRAMARAGITGADVAVVEAHGTGTRLDDPVEASALITATYKSGSSG  
PVLGYSVSNIGHAQAAGVAGVILKVLGERGVVPMLCRGRSGILDMSGSRIELA  
DGVEMSPAADGVRAAGVSAFGVGTNAHVIIAEPPEPVPQPRMLPATGVVAVVL  
SARTGALIRQAQGRADHLIAHPGILAPDAGSMTMARPOHPEERAAVLAADTTAAVHR  
LRVADGAVVPGVTTGASDGGSVFVFPQAGQWGMARELLPVVTRAESIACEDAVIL  
SEVAGFSVEVLERPDAPSLERVAVQPVLFVAVMSILARLMRACGAPVSATVHSQG  
EIIAAVAGALSLDEGMKVVARRSRDAVAVAGRSMLSVRGSRDVEKLLADDTWTR  
LEVAANGPDVAVVAGDOAREFLECEVGIARALPVVYASHTAHVALEVEBELVQ  
LAACITPRRAVEPFESTLTGDFLTFELDEAGVYRNLRHVPEFSAVALLODQYATF  
IEVSPHPIVLSVOETLDDASDAVYITLERDAGDARFLTALADTRVAVDMWA  
VLGRAGLVLDLPGYFQGRFVLLDRTTPRDELDGKTRVDVTEVPSSEPRALGRML  
VLPBEGHEEDWTVRSALAEAGABPEVTVGGVGLVGDCKGVSLALRBDGAVQTL  
VLVRELDABEGIDAPLMTVTFGADVAGSPVAPDQAKLGLQVASLEREGPGLVLDL  
PHMPDELRGLTAVLAGSEDOVAVRADAVARARLSAPAVTATSEYAVPGCTIIVTGG  
TAGGLAGVMLAGRGAHLLVSRGDDTEGVDLTELRLGARVSVHACVDSRE  
PVRILVHGLTROGVVGVVHAGLPOOVAILNDMDEALFDRVAVAKAGAGVHLDLCS  
DAELFLPSSGAGVWGSARQAYAAKNAPFLDAPARHRGRCIPIATSVAKGIMAGGNT  
GDEBARSFLARSRGAVAMPVRPALAALDRVLASGETAVVTDVMPAPAESYTAARPP  
LILDIRVTTAPSRBAGEPTEBESLDRLAGLPAESRTAEVLVTRTSTAVTLGHDPKAV  
RATPPEKELGDSLAAYLRMLNAATGILRPLSTLVFHPNPAASVAGLDELCTEYR  
GEAPSAIAGDALRALPVPVATERBEELVGLLEHMLAALRPVAAQAADASGICAMPSSG  
DLGRAGVDELEALGRELDG"

/note="DBS2"  
/codon\_start=1  
/transl\_table=11  
/product="EryAII"  
/protein\_id="AAV51821.1"  
/db\_xref="GI:55419416"  
/translation="MTDSERVAAVYLRATLIDLRANQRIELESDPATISMARCLPG  
GVNTPRLMELAREGETLSGFTLDGMWDLARLHHDPDPNPGTSYVDKGFLLDDAAF  
DAFFGVSPREAAAMPQOQLLETTEWELVENGIDIPSRGTRGATVLAALAEVCAAE  
DTPAAEDBEQSVTVAPASRGISITWMLDEPSSISVDPRACSSLYALALVAESLAK  
GESMAVVGGAAYMATRGVAVDSRQALADRSRARGADGFGFSRGCVTTLIER  
LSRRKNGEYTLAVRGSALNDQASGSLASBEGPQORVIRIOLAESCGLEPDDVAV  
EAGGTALGDPIEANAALDITGRDDADPLWGSYKSNIGHTQAAGVTLKTVL  
VILNGBELPATLHRETEPTHVDSGGVALLAQNQPMRGRTERRAAVASAGISGTNAH  
VIVEAREPHERHETTHADGRPVLVVSARSATLAPQAAIAELRETFDADLAGVIG  
LATTPRAHEIRAAVSTREAREGRILEAAGATVDAVEGTVAGVAVVLEFPPQ  
GSMWAGMALLSSPVFAKTIKACESMAPMQDMKYSDVLRQAPARGLDRYDVVDP  
VLAVWVSLAEILKRSIGVEPAAVVGHSQGEIILAAHYAGALLTEDAKLTVGRSLKRS  
LSGEGMAAVALGEAAVRELRPMQRLSVAANVPERSVVSGEPALRAFSEDCAAE  
GIVRPIDVDYASHSPQIERVREBELLETTGDIAPRPVATFHSITVESRSMQDELAR  
YVRNLETVRFADAVTRLAESGVDLIEVSPHVVVAVBEAVEEADGAEADVAVGS  
LHRDGDGLSALSMATAHVSQVDIRMDVALPQAPALPTYEPFORKRYMLQAPAPAA  
ASDELAHYVSWTPIEKESGNTLGGDMLVYTPILSPWMTMLCEAIVNGGRALRCEVD  
TSASRTEMAQVADAGTGFRTVSLSSDSACRPVPAWAGILTTIYQALGDAGVLA  
PVNCTQGAVRTPADDDLARPAQTTAHGFQVAGLELPGMGVGVDPSSVDDALRG  
LVNVLARGAEHDLAVRDRGLHGRVVRASLLQSGGSRWTPHGTVLVGSAASPVGQ  
LVMWLDRGAREVLVLAGACGDLLAAVEBAGASAVCAQDAAALREALGDEPTTALV  
HAGTLNFGSISVAPREFAETIAAKTALLAVDEVYGDRAVEREYVCSGAVIMGA  
GMAAYAGSXYDLALAEHRARGRSCVSWTPTMALPGCAVDDQGYLREGRILSADR  
AMTWERVVLAVGVSVAADVDPVLSBGRPATRPTALPAILAGRQQAAREPDSGT  
KEBAPALRSGESPQOENLELVANAAVLAGHSAAEIVNRAFSELGLDLSLAML  
RKLSLSTGRILRLPSLFEHPTVTLAQHILRALVGDADAAARVAGADBSERIALY  
GIGRFPFGGSGPEOLMRVLAEGANLTGPRAGVREIRGELVDPNPTGTSYVDKGG  
FLDDADPDDGFGFITPRREALNDPOORLLETAMBEAVERAAGIDPDLRTGDCVFG  
MNCQSTYQQLAGAEERVDYQIGKNSAYTSGSLIAYFGMEGVALTYDTRACSSSLVSI  
HLMAQILRSGESDLAAGGVTVNSDPTTFVSTQKSLASDGRKAFSARAADPALS  
GVVALVLEPSSRRANQVLAALRGSANVDQASNGLANPESQORVIRIOLAAASG  
VPADLVVVAHSGTGETLGPIDBAGALIAFYQGDORPELFGKRTIHTQOAAAG  
GADIKVVLAMGHGPRELHDEBEPHIDMGAVREYREVPMPAGRPRAAGVSFG  
VSGTNAHVITYERAPABQEAARTGRGPLPTPLSGRSFVVAQAALVALEHLDPELIL  
TDAANTLATGRAPFVRAAVTGDPRAGVCALDMLAEGRSADAVALPYSAPRPVTV  
FPQGAQWGMAMADLLESSVFPABSSRCBAWLSPTHDMKLVLDVVRDGGPDRERVD  
VLDPVLESVNSLAELMRAGVTPAAVVAHGSQEEIAAHVAGALSLEAAKVVALARQ  
VLEBILDQCGVAGSARDELTVLAWMDGRVQVAAVNGCSTVAPPTAELEJFPE  
ABAREMKEPRLIARVYASHPEVRIEDRLAEIGTIRAVGSLVLSHTVAGEVIDTSA  
MDASVYRNLRRPVLPEQAYRGVGEQFFQFVSPHVLVMEVETAAHRAAEVTCV  
PTLRREDSGHEFLKRLIRAHVHSGEADLRPAVAGSPALPTTPFBRQFAPRPNHR  
ADVASLAGEVRAEHPULLAAVDVPHGDAVFTGRISTDEQWMLAEHVGATLVGSLV  
VLDLAAAGEVNGVPLELAVLQRPVLVLAGAGALLRMSGAPADESGRTTIVHAAEDVA  
DLDAQMSOHAFTIAGVAAAGPDTQMPEDAVRILPDDHVLGLAEOGYEYGPSRQ  
ALBAAMRKDDSVAAEVSIAADESGVAFHVLVLDVAQTLISGLALGSGGKGLPRAMT  
VTLHAGSATSVRYVATPRAGDAALARTDPDAGHLVAVVDSLVVRSNGEKKQBPBRG  
EGEHLAALDWRLEPSTGTGVNADASDLDAVLSGSEPPEDAVLVRTEPBGDPRAA  
RHEGLMAALVWRMLRQEBELPGATLVITAGAVTBDDDSVPEEGAAMMGVIRCAQA  
ESDPRFVLTLDPEPGLPAVNPDPQALGCDVFFVLEPPLPBLATLPLAGTQRLV  
GDGADISGAEVAPDVQOPPRAGVREYDVATGTFNFDVLLALQMTYPOKMDTEAAG  
VTVRAGVDVDAVFAFGDRLGLFGQAFPIAVTDBRLASFVLLPVDGSDDAALVPAVTT  
AHYALHDLAIGRAGQSLTHAAGVGMVAALABRGAVALVATAGRAKIGTILPAILTL  
DDEHTIASSRTEGPKRREKTGGGQVDTANSITGELLSDSADLDAEDGVFVEGKTD  
LIRQADPRGKVVLPDDEADDDIGETLIRGVVGLLGRGALDRLPVSEMEISASRALQ  
HMSGRGVAGVLPDOPVAPVDDGVLTIGETVGLGRLLARHLVTEHVRMLLVSRG  
ADARSGELAEIEIDLASAEIIACPDADDAISALDGLPRLPTVLAAGVAVAGLGL  
VTSIDPEAVSOVRAKYDAMNIELNARLADLSPFLVFSASAVSLAAGPGGVAAANES  
SUNALALARTRCRIPAKALCWMGLMAQASBMTSGIGDPIARTQVALPTEERALLPDS  
LIRGGEVVPFLSIRSLARBAEPVBLRKQVRAFLPDAQOAGLAEGRVTRDLRAGRE  
SDQVAGLAEIVLRSALAAVSGSADQIPERKARFDQGLFDSIAAEILNRVIGTAVGVL  
PSTLVFHPPTPLAAEHLRDLFAASPAVIVIGDLIDELERALELSADEGHDVGORL  
ESLRRMNSRADPSTSAISEDASDELSMLDPRGGGSDL"

gene  
CDS

gene  
CDS



RKLSASTGRLPASLIVDPHTVATLALQHLRAALVGDADQAAVRVGAASESPILAV  
GIGRPFPGIGSPQOLMRVLAEANLITTPGADRMGIMIGRLVDPDNPGRSYVDKGG  
ETIDADPDPCPGCTPRRLALANDPOORLIMLKTAMRVERHAGIDPDOLRETDYFPG  
MNGOSTRQQLAGBERVDGTCGLNSAVLSGRATTFPGHGLTDTTCSSSVIGI  
HLMAOLRKBESCIALAGVTVMSDPTTFVDFSTORSLADGCKASADALUSG  
GVALLVLEPLSRARANGHOVLAVLRGAVQDQNGSLAANPGOSORVIROLAAAG  
VPADVAVVEAHGTGLDPIEAGALITATGODRDRPLTGSYKTIHGTQAAAGAA  
GIVKIVVAMRGMPLRSLHDELSPHIDMSGAVLEERVPMPAGSPRAGVSG  
VSGTNATVIVEABARBOEARTEGRCPLTSGSEAVVAOARALAEHLRDPDEL  
TDAANTLATGRAPDVAAVATGDRACVCAELDLADGRSADAVALVTSAPRKPVIY  
FPQGAQWVGAKARDLLESSVEFASMSRCABLSPTDMKILDVDRDGDQDHERVD  
VLQVPLSIVMSLAEWRAGVTPAAVVGSGEIAAHAVAGALSLBAAKVVALRSQ  
VIRELDDQSGVSVGASRDELTVLARMGCVAAVAVNGPQTSVAGPTAEHLDFEFA  
AABEMKPRILAVVYASHPREVARIERLALCTITAVNGSVLHSTVAGEVDTVA  
MDASVYRNLRAPVLFRQAVRGVYGEPTDFEVSPPHVLMAVERAHRAAGEVTV  
PTLRBEGSPHEPLRMLRHHVGVADLIRAVANGRPALPTTFEHCQFVRPHRP  
ADVSLVGRGAHPULLLAADVPEHGAVPTGRLSTDBQPMPLSHVVGRTLVGSLV  
VDLALAGEVDGLPVLBELVLRPLVLAGAGALRMSGVAPDESGRRTIVHAADVA  
DLADQNSOAHGTGLAQVAGAPDTEQWPEDAVRIPLDHYDGLAEQGYEPPSQ  
ALRAAMRKDSVYAVESVIADEBEGVAFHPVLDAVQTLISGALGEEGGGLPEAMT  
VTTLASGATSVRVATPAGADAMALRVTPAGHCVATVYDLSVYVSTGKNOPEPRGG  
EGELHADMVLAEPGSTRGVVADAADDAVLSGSEPRPAVTVRTEPSDDPRAA  
RHGVLMAALVRRMLEQBELPGATLVATSGAVVSDDSVPSGAALAMGVIRCAQA  
ESPDRETVLDTDAEPGMPLPVPDNPOLARGDVFPRLSPLASALTLPAGTORLVP  
GDGAIDSVAFEPADVPOLRAGEVRVDVRAVGVFRDVLALGMYPOLKDMTEAAG  
VTVAVDPDVAFAPGDRVGLFQCAPAPLAVTDRRLARVDPGSDDAAVP IAYTT  
AHVALHDLAGRAQSVLIRHAAAGVGMALVLRBAGAVTLAPGAGKCTTALAL  
DDEHIASRRETFARKRERTKRGVDTVNSLTGELIDESADLAEQCFYVEBKTD  
LRADGDFRGRAYAPDLGEAGDRLGEILREVGLLAGEEDLRLVPSAMELISLAALQ  
HMSRGRHVGKVLTOPAPVDPDGLVLTGCTGTGLRLRLVTEHGVRLILVSRG  
ADAPGDELRABEIDLGAARILAACDTADRALSALDGLPRPLTGVHAAGVADL  
VTSIDEPABOVULAKVDAANLHETLANTGLSPVLFSSASVLAGGQGVAAAE  
SINLALARTRGIPARAKLGMGLMAQNSMTSGCDRIARCVGVALPTREPLALPDA  
LRKGEVVPFPLSIRSAKRAREFVSLRGMVRAKLRAGQAEAGRVNVDRLAGRES  
SDQVAGLAEIVSRHAAAVSGYSDQUPERRAFDOLGPDLSLAVALERLSTAGVAG  
PSLTFDHPRLTAAVEHLRDLFAASPAVDITGDRIDELRKLALSLAEDGHDVQORL  
BSLRMRSPRADPSTSAISEDSDELFSMLQRFQGGSDL"

misc\_feature  
/gene="erya"  
/product="beta-ketacyl synthase"  
1471..2763  
/gene="erya"  
/product="acyltransferase"  
4261..4455  
/gene="erya"  
/product="acylcarrier protein"  
4513..5787  
/gene="erya"  
/product="beta-ketacyl synthase"  
5839..7116  
/gene="erya"  
/product="acyltransferase"  
7117..7563  
/gene="erya"  
/product="dehydratase"  
8743..9234  
/gene="erya"  
/product="enoylreductase"  
9418..10164  
/gene="erya"  
/product="ketoreductase"  
10315..10509  
/gene="erya"  
/product="acylcarrier protein"  
10750..20268  
/gene="erya"  
/note="ORF A"  
/codon\_start=1  
/transl\_table=11  
/product="6-deoxyerythronolide B synthase III"  
/protein\_id="CAA44449.1"  
/db\_xref="GI:46979"  
/db\_xref="GOA:Q03133"

/db\_xref="UniProt/Swiss-Prot:Q03133"  
/translation="MSGDNQMTREKRLRYLRKRVTEIDSVTALREVERHAGPILAV  
GMACRPGVGVSGDFEYFSGGDAIAELAPRAGMSPDDALGQALAAQFADAF  
FGISPRBALANDPOORIMLEISWEALREKDVVLSRGATGVGTGTVDYGRPDE  
APRVLGYVTGTVASGRVAVCIAGLSRPMNTDVCSSGLTALHMEISLRDSC  
GLALAGVTVMSPGAFTFRSGGLAADRCRPSFKADQDFGLAGACTVLCORLSA  
ARREGPVLAVVAVNDQANGNLTABEGQKQNTIRLALNAGVRADVTVREAH  
GTGTRLDPILEVALLSTYGAEDPDPLMTIGSVKQNTIHTQAAAGVAKVTLAR  
HGMPTLHEDPSPOIEMPLVGSVQSAWPAERPRAAVSGFISGIGTAAVTV  
EAPRADPERPADSGVPLVLSGDBQRAOAGLADHLAEPRSLADTGTTLT  
RANMERAVVGVDRDBALAGLAVADRILADTTACCGNSPRVAVMPGGQAMQ  
GMARDLRESVADSIRODERLAPRVDLTLDSGARPLRVRVVQALPAAVMS  
LAALESFHEPAAVVGSGEILAAHVALTLLEAAKVAARSLRLRLQGGMA  
SFGIETEOAERIGRPAGALISIVSGSVVAVVAGSGEILAECEAEARKARI  
PVYASHSPVESLRESLTELAGISPVSDAVVASTTQCPDITAMDAVYANIR  
BOYRPODARTQLAEAGDVAVSVPHPLVYEAUTDPLVDPAQCVGTGLRDDM  
GLADFTALGEAVACQVYVMSPARPADRVPELVPYDQYMLPIPGGABRDDDM  
RYOVVREABMSASLAGVLLTVGRPVEBSLDAISGLBQGAIVLTCDSYRSSTI  
GTALAEADTDLASTVGVAAPHEBAVPSLDLALVOLQAAGVAVELVTRNAVQ  
ADELVDPAQAMGGLGRVVGIGPQMGGLVDLVADAASISLAVLADPGEBOV  
AIBADGIVKRLVLPARARTHLEPLAAGTVLVTGGGIGAHILAAVLAEGASHVL  
LGRGADAQASLRESLTALGTGVTIACDVADRARLEAVLAABAAEGRTVSAVH  
AAVYSTPDLDTTEAFETIADYKRTGYNDLDELCPDLDAVTLFSGNALVMSPGIA  
SYAANAFLDGFAPAAASBGAPVTSIANGMLAQOMAGDGEDELAEQIRANDPDA  
VBEHLITLDHGQTSVSVMDRRPVELFTAAHRPLPDSIAGARARQSBQGLA  
ORLALLCDREHREHLIRAVVAAVLEGDDAIDRDAFPDGLFDSMTVDLNR  
LAAVTVREAAVTVFDEHTITRLADHLETVLGAALAEQAPALVREVPDADPILAI  
VGMACRPGVNVHMPGLMERIVGGQAVTEMPTDRGMDLALPDPORGSTYSRIG  
AFLDGAADPAAFPGLSPRALANDPOORVLTETWELFENAGIDHISVAGDQVTL  
GAAYVQGGVAVPEDEBEGYLLGNSSAVSGRAVVLGLEGVLTDTACSSSLVAL  
HSQGLLRDCCGLAVAGVSVNAGVPEVTEFRQGLAAGLADGCKASABADGGLBE  
GVAVQOLRSDPABEGROVLVAVGASAINQOATGGLAABGVAQOQVIRKMAAA  
GIGADVAVVEAHGTGLDPIEAGALITATGOSRSGSPVLTGSYKTIHGAQAA  
GAGVQIKVIVGAHGTGLRDLPEASALTLATGYSRSGSPVLTGSYKTIHGAQAA  
SAGVSGTNAHVITAEPRPRLPEPRPVVLAANSPVTLKARTTALAAORLLE  
SAVDSPVTLALASALTGSAHLEKRALAGHSBRLAQGLRAYABSVAPGATTCRA  
SAGGVVFPQQAQOMEGMARGLISVFPVLEABCAVLAEBVAFASVELRBPRA  
PSLERVDVQVPLFSVWVSLARLMGACGSPVAVTGHSGOEIAAAVAGVLTLEDPA  
VVLRAKRLALAGKGMVLAABEGARALIAEMERISVAAVNSGVSVVSGDPPRA  
LAEI VACDEBGVRAKTLIPDVYASHSHSVEIRITLADIIGISARAAIPLYETLNG  
ERKMGPRTYDNLKSGVRDEAVASQSPGHATFYMSHPHTTAAVOITAAVAVI  
GSLHRDTAEHLIAELARAVHGVAVDMVRVPPAPVPLPNTPFEQRTILAEVVD  
QLADSRVYDVRPLATTPVDLREGFLVHSGAPBELTAVBEKAVGVVVAASDRALAA  
ALREVPGEVAGVSVHTGAANALALHOSLGAGRAJMLVTSRAVLAIGSEPRVAD  
AMVWGLRVLGSETPBERMGGLVDLPAPARPDGGAFAVACGADHEDLOVALRDAJRA  
RUVYARLGTRESSEWPAAGTALVGTGGVAGRAHRLAAGVDELDLVSRGDAPRA  
AAELBAELVALGPKTTTITACDVADRBOLSTLBEILROGKRPVVVHTACVPSERPLH  
RIGLESVCAKAVTGARLDLBCPDATFTVLFSGAGVMSANLAIISANAITYDALLA  
HRRABORAAITSVAMGAMAGBMAVGLBGLTRRGLRPMAPDRAIRALHQAANDGTC  
VSIADVMEAPAVGFTAPRPLDELITVAVAGVAVPVAAPARPMTSQELLEPTHSH  
VVALTGHSSPDVAGDOPFTLEIGDLSLAVGLRVLQQAATGLALPATLVFHPVRLH  
ADHTGOOLDGSTRBARBASALRDGYRQAGSRSTYLDLAGSDFREHSDGSGFS  
LDLVMDADGPREVTVICAGTAIISGHEHTRLAGALRGTAIPYAVQPOGYEBOEPPL  
SSMAVAVAQADAVITRQGDPRFVVAHSGALMAVALATELRLRGHPRKAVVLTIDY  
PQGHQDAMNAMLBEITATLTPDRBTRVNDTRLTALGAYDRLTQWRRETEGLPTLLVS  
AGEBPMGPMPDSDSMPTWPFSHDTVAVPGDHFVQEHADAIARHIDMLGNS"

misc\_feature  
/gene="erya"  
10867..12111

Query Match 44.9%; Score 269.2; DB 1; Length 20444;  
Best Local Similarity 68.0%; Pred. No. 8.5e-17;  
Matches 408; Conservative 0; Mismatches 183; Indels 9; Gaps 2;

QY 1 CCGGCTGCGCGAGCGCTACACCCGCGGGGTGGAGCTCGACTGCGCGTACCGCAGTGGGTG 60  
Db 13337 CCGGCTGCGCGAGCGCTACACCGCGGTGGAGCTCGACTGAGCCCGCCTTCGCG 13396  
QY 61 AGGACGCCCGATGCACTGCGCGGTCTACCGGTTCACAGCAGACAGACTTGTGCTCCCGG 120  
Db 13397 ACGCGCGCGCGATGCACTGCGCGGTCTACCGGTTCACAGCAGCAGACTTGTGCTCCCG 13453  
QY 121 TCCCGTGGCGCGGGTCCCGACACCGCGCAGAGTGGCGTTACCACTGCGCTCCGCTGGCAC 180

Query Match 40.3%; Score 242; DB 1; Length 9769;  
Best Local Similarity 64.9%; Pred. No. 3, 5e-14;  
Matches 409; Conservative 0; Mismatches 185; Indels 36; Gaps 2;

mat\_peptide  
57. .9590  
/gene="erya"  
/product="6-deoxyerythronolide B synthase"

ORIGIN

1 CCGCGCTCCGCGAGGGGTACACCCGGGGGGTGAAGTGCAGCTGCGCTACCGCAGTGGGCG 60  
2638 CCGCGCTCCGCGAGGGGTACCGCGAGGGGGTGAAGGTGCAGCTGAGACCCCGCTTCGCGG 2697

QY AGGAGCGCCGCGTGCAGCTGCGCGGTTCACCCGTTCCAGACAGACGACTTCTGGCTCCCGG 120  
Db AGCGGGGGCGCGTGCAGCTGCGCGCTTACCGGTTCCAGGCGCAGCTGCGCGCTGTCACCGCT 2757

QY TCCCGCTG-----GCGCGGCTCCCGACACCGGCG 150





terminator  
10815..10861  
promoter  
10882..10899  
misc\_feature  
10902..10926  
gene  
10933..21677  
RBS  
10933..10938  
CDS  
10947..21677

GEAPSLAAGLDALEALPEVPATEREELVORLERMIALALPVAQAADASGTGANPSGD  
DLGEAGVDELLEAGRELDGNSHHHHH"  
/note="T7 terminator"  
/note="T7 promoter"  
/note="lac operator"  
/gene="eryAII (morphed)"  
/gene="eryAII (morphed)"  
/note="similar to Saccharopolyspora erythraea EryAII of  
GenBank Accession Number A661566; contains His-tag"  
/codon\_start=1  
/transl\_table=11  
/product="EryAII (morphed)"  
/protein\_id="AAV39551.1"  
/db\_xref="GI:54778586"  
/translation="MTDSKVAEYLRATITDLPAAQRITELSDPAIVSMARLPG  
GVNTPQRLMELRGGSTLSGFTDRGMWDLRHHHPDNPSTSYDKGFPLDADAF  
DAEPFGVSPREAAAMPQORLLETSMELVKNAGIDPHSGRTATGTVFLGAVKGE  
DTAAAEVVEGYSTGVAPAVASGRISTYMLGSPSISVDACSSSLVALHAYELRK  
GESSMAVGGAAVWATPGVFPFSDROBALADGRSKAFGAGDGSGEVTUILLER  
LSRRRNHGEVLAIVRGSALNODGASNGISAPSGPQORVITRQALSCEGEPDVAV  
EAGGTCTALADPIEANALDITGDRDADPLWGSYKSIIGTQAAAGTGLKTVL  
ALNGBLIPATLHVEPTPHVMSGVALLAGNQPWMRGRTRRAASAGISGTNH  
VIVEBAPEHRETTADHGRPVLVVSABSTALRQAAQIAELIEPDLADGIVG  
LATRAHERAAVASTREAVRGLREIAAGATADAVEGTVRVDGRVFLPFG  
GQWAGKGAELISSPVPAKIRACDSMA PMQDKVSDVLRQAPGRGLRDVIVOP  
VLEAVWVALELRSGVPEPAVVGHSQSIIAAHVAGALTIEDAKUIVGRRLKAS  
LSEGGMAVALGBAAVERKLEPMQDLSVAAVNGPSSVSGPGALRAFSEDCAB  
GIVRDI DVYASHSPQIERVERBELLETTDIAPRPARVTHSVESRNDGELAR  
WYRNRLSETYRFADAVTRLAESGDYATIEVSPHVVAVEBAVEADGADAVVGS  
LHRDQELSAFLRSMATAHVSGVDIRMDVALPGAAPALPTYPEFORKRLQAPPA  
ASDELAVRVSMTPIEKPSGNLGMWLVVPLISPEWTEMLCEAINNGRLARCVD  
TSASRTEMAQVAAQGTGPRGVLSLSSDSACRPGVPAPAGVGLITIVOLGACVDA  
PVNCTQGAVRTPADDDLAPQOTTAHGFPAOVAGLEIPGWSGVVDIPESVDALRL  
LVAVLGGGAEHLAVRDRHGRVVRASLPQSGRSWTPPGTVLDEAPFVGL  
LVWMLADRGAEHLVLAGACGDOLLAVERGASAVVACADQALRELDDEPVALY  
HAGTNNFSGISSEVAPPEFAETIAKTAALLAVDEVLDGAVREVEYCSVAGIWGGA  
GMAVVAAGSAVTLAELHHRABGRSCSVATPMWALBGA VDDGTYLREERLSLSDR  
AMRTERVLAAGFVAVADVDPVTSSEGPATRTPLFELAGGQDAEPDSEPT  
GEAPQARLSTGSPDOENLLELVANANAVALGESAELINVRASELIGDLSNAML  
RKLSASTGLRIPASLVFHPHTVALAQLRILSSPADDAARVAVGADSEPIALV  
GICRPFGGISGPEOLMRVLAEGANLITGPPARGMDIGLHYDPDNPSTSYVDKG  
FLTDADPDGPGFITPRBALAMPORMLETWAMEVBERAGIDPALRGTDGVPVG  
MNGOSTYQMLAGABRYDGYQIGNSASVLSGRITATTCMBGALTYTDRCSSLVGI  
HLMQALRRECSLALAGVTVMSDPTTFVSTQRLGADGCKAVASADSPALSE  
GVALVLELSPRANGHOVLAVRGSAAVODGASNGLAANPNSORVTRQALASG  
VPADVDVVAHGCTEIGELBAGALILATYGODRPLBLAGVKNIGRPAQAGAA  
GTVLKVLAHMGHGLPSLHDELSPHIDMSAGVAVREVPMPGPRGATGAVSNG  
VSGTNAHVVERA PAQOEAARTGRPLPVLSGRSGAVVAQAALAEHLRDPBGL  
TDAATLTATGRAPFVRAAVLGDPRAGVCAELDAEGRSADAVPTTSAPKPKVIV  
PFGQAQWGMWMDLLESSSEVFAESMSRCEALSPHMDLILVNGDGDPEHERD  
VLOPVLFSINVASLAEIRAHGVTTPAAVVGSGEILAAHVAAGLSLEAAKVALAL  
VLELDDQGMVSGARDELLETVALRMDRVAIVAAVNGGTGVAGPTELDPEFAB  
AABREKMPRIIVAVASHSPVARIERLAEGLTITAVGSLPSTVAGVETLVCV  
MDASVYRNLRPVLPEQAVRGLVEQGFDPVAVSPHPLVLAVERTAHGALEIVCV  
PTTIRREGSGHEFLRNILRAHVHGVGADLRPAVAGRPALPTYPPEHQFWRPRRP  
ADVLSGVRGAERHPLLLAADVFGHGAFTGLSTDEQWMLAEHVAGGTVLPGSVL  
VLTALAGBDVGLPVLBELVLRPLVLAAGALLRMSVGA PDSGRTTIDVAABDV  
DLADQMSQHATGLAAGVAGPRDEHPEDAVRIPLDDHDTGLAEQGEYGPSPQ  
ALBAWRKODSVAAVSIAADEBGVAFHPVLLDAVAGTSLGALBEGGKLEPANNQ  
VTLHAGATSVRVATPAGADAMALVTPDAGHVAITVNSIVRSGCKEKOBEPRG  
EGELHALDRTYRLABPGSTGRVVAADSDDAVLRSGEPEDAVLRYRBPBGDPRAA  
RHGVNMAALVVRMLQEBELPGATLVATSGAVTSDSDSVPEPAAAMGVIRCAQA  
ESPDRPLVLTDAEPGLPAVPDNPQALRGDGVFRLSLPASPALTACTORLVP  
GDGALDSVLEPAVDVQPLRAGEVRDVATGVPNDVPLALPAPOKRDMGTEAG  
VTVAVGPDVDAFAPGRVITGLPQGAFAPIAVTQTHRLIARVPDMSADAAVAIATTT  
AHYALMDIAGLRAGQSVLTIAAGCGVMAAVALARAGAEVLAITAGPARKGITRALGL

DDHIASRETFARKFRRTGRGVVDVVLNSLTGELLESADLLAEDVFEWEMKTD  
LRADGPRGRVAPFDLGEAGDRLGELILEVEVLLAGELDRIPVSMELGASPAALQ  
HMRGRGVKGLVITOPAPVDPDGTUJITGGTGTGLLAHLVTHBGRVLLVSRGG  
ADA PSDDELRAIREDIAGAEIACDTPARDALSALLDGI.PRLTJGVVHAAGVADBL  
VTSIDPRAVEQVTRAKVADAAMNHEITANTGLSPFVLFPSSAASVLAGPQGVTAANE  
SLNALALRRTRGLPAKALGWMGMAQASSENTSGLGRITATGVALPETERALLFDSA  
LRGGGVVPEPLSINRSALRRAPVPLRGMVAKIRACQAAEAGAPNVVRLRAGRE  
SDQVAGLAELVRSAAAVSGYADOLPERKAKIDGLGFDLSLAAYELNRLGTAGVGL  
PSTLVDPDPLPLVAARHLDRLSASPAVDIGRDLBELKELALSADGHDVQGL  
BSLRRMNSRRADAPSTSAISBASDELPSMLDQRFSGGEDLGNSHHHHH"  
/note="T7 terminator"  
/note="T7 promoter"  
/note="lac operator"  
/gene="eryAII (morphed)"

Query Match	31.5%	Score 189;	DB 11;	Length 31531;
Best Local Similarity	59.5%	Pred. No. 1.6e-09;		
Matches 359;	Conservative 0;	Mismatches 235;	Indels 9;	Gaps 2;
QY	1	CCGCGCTCGCCGAGCGGTACACCCGGGGGTGAGGTGCACTGGCTACCGCAGTGGTG	60	
DB	24455	CCGCAATTAAGCGGAAGCCTATGCCAGGGGTGAGGTGATGGTACCTGCTTTTCGG	24514	
QY	61	AGGAGCGCCCGGTGCACTGCGCGGTCTACCGCTTCAACGACAGAACTTGTGCTCCGG	120	
DB	24515	ATCCCGCCCAAGTGAATTAACGATGATCGTTTACGCTCAAGCTTACGCGTGCAGA	24574	
QY	121	TTC---CCCTGGGCGGGTCCCGGACACCGGCGACAGTGGCGTTACCACTGGCGCTGGC	177	
DB	24575	TTCGACAGGTGGCGGGCTCTGAGCAAAATATATATATGCGTTACAGTGGCTTGGC	24634	
QY	178	ACCCCGTCGACCTCGGGCGGTCTCCCTGGCGCGGACGGGTCTGTGTGACCGGAGCG	237	
DB	24635	GTTAAACGGAATGGAGTCTGCTCCTCGCGGTGCTGCTGTGTAACCGGCCCG	24694	
QY	238	CAGTACCCCGGCTGAGACGAGTGGTCCGACGCGCTGAAACAGCGGGCGGACCG	297	
DB	24695	GTTACCATCTGAGTGTCCGATGCGCATCGGTCAGGGCTGAGACAGTGGGGCAACGG	24754	
QY	298	TGCTGTGTGACACCGGCACTCGCGCGCGGATCGGCGCGGCACTGACCGCGTGAACG	357	
DB	24755	TTTTGAATGACGACGTCGAAAGCGCTTCCACGATCGGACAGGGTGGAAAGCTGTGATA	24814	
QY	358	GCAACGCGCTGTCACATGTGTCTCTGTGCTGCGCGCGGAGGGCGGTGTGACAG	417	
DB	24815	CTGATGGCTGAGACCGTATGATCGCTGTTAAACCGTG-----ATGGCAGAGCTTGC	24868	
QY	418	ACCCGAGCTGACACCTCGCTGTGTCAGAGCGCTCGGCGGACCGGATCGACGTCC	477	
DB	24869	ATCGAGGTCTGATGCTGTGGCTTTGGTGAAGCCCTTAAGGTGCTGGGTGGAAGCAC	24928	
QY	478	CCTGTGGCTGTGACACGAGGACCGCGCGCGCTGACCGTTCGAGACAGCTGATCCGG	537	
DB	24929	CGCTGTGGGTCTCGAACCCGTAAGCTGTCTCAGGTTCTGATGTGTGAGCTGTGATCTTG	24988	
QY	538	CCGAGCGCATGTGCGTGGGTGCTGGGCGGGGTGTGGGCGTGGAGTCCCCCGCGGGTGG	597	
DB	24989	CCCAAGCATGTGTGGGCGGGGTGGGCGGGGTGTGGTATCGAACACCGGGTGTGGG	25048	
QY	598	GTG 600		
DB	25049	GCG 25051		
RESULT 10				
AB088224/c				
LOCUS	AB088224	210614 bp	DNA	linear BCT 11-JUN-2003
DEFINITION	Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.			
ACCESSION	AB088224			



VERSION	AB088224.1	GI:30698345
KEYWORDS		
SOURCE	Streptomyces rochei	
ORGANISM	Streptomyces rochei	
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
	Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	1	
AUTHORS	Kinashi, H., Fujii, S., Hatani, A., Kurokawa, T. and Shikawa, H.,	
TITLE	Physical mapping of the linear plasmid pSLA2-L and localization of	
JOURNAL	the eryAI and actI homologs	
REFERENCE	Biosci. Biotech. Biochem. 62, 1892-1897 (1998)	
AUTHORS	2	
JOURNAL	Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H.,	
REFERENCE	Nishi, O. and Kinashi, H.	
AUTHORS	Identification of two polyketide synthase gene clusters on the	
TITLE	linear plasmid pSLA2-L in Streptomyces rochei	
JOURNAL	Gene 246 (1-2), 123-131 (2000)	
REFERENCE	10767533	
AUTHORS	3	
JOURNAL	Hirateu, K., Mochizuki, S. and Kinashi, H.	
REFERENCE	Cloning and analysis of the replication origin and the telomeres of	
AUTHORS	the large linear plasmid pSLA2-L in Streptomyces rochei	
TITLE	Mol. Gen. Genet. 263 (6), 1015-1021 (2000)	
JOURNAL	10954087	
REFERENCE	4	
AUTHORS	Mochizuki, S., Hirateu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K.	
TITLE	and Kinashi, H.	
JOURNAL	The large linear plasmid pSLA2-L of Streptomyces rochei has an	
REFERENCE	unusually condensed gene organization for secondary metabolism	
AUTHORS	Mol. Microbiol. 48 (6), 1501-1510 (2003)	
TITLE	12791134	
JOURNAL	5 (bases 1 to 210614)	
REFERENCE	Mochizuki, S., Hirateu, K. and Kinashi, H.	
AUTHORS	Direct Submission	
TITLE	Submitted (15-JUN-2002) Haruyasu Kinashi, Hiroshima University,	
JOURNAL	Department of Molecular Biotechnology, Graduate School of Advanced	
REFERENCE	Sciences of Matter; 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima	
AUTHORS	739-8530, Japan (E-mail: kinashi@hiroshima-u.ac.jp,	
TITLE	Tel:81-824-24-7869, Fax:81-824-24-7869)	
JOURNAL	The nucleotide sequence has been determined by using restriction	
REFERENCE	fragments and nested deletion fragments of the ordered cosmid	
AUTHORS	library of pSLA2-L.	
TITLE	pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries	
JOURNAL	143 ORFs.	
REFERENCE	Gene prediction was based on the unique codon usage in Streptomyces	
AUTHORS	(Bibb et al., Gene 30:157-66 (1984)) using the Frameplot program of	
TITLE	Ishikawa and Horita (FEMS Microbiol Lett 174: 251-253 (1999)) as	
JOURNAL	implemented at	
REFERENCE	http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl. Where possible we	
AUTHORS	chose an initiation codon (atg, gtc, ctg or ttg) which is preceded	
TITLE	by an upstream ribosome binding site sequence (optimally 5-13 bp	
JOURNAL	before the initiation codon). If this could not be identified we	
REFERENCE	chose the most upstream initiation codon.	
AUTHORS	Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes	
TITLE	(three PKS gene clusters for lankacidin, lankamycin and an unknown	
JOURNAL	type II polyketide, a cardenoid biosynthetic gene cluster, many	
REFERENCE	regulatory genes and others). The range of each biosynthetic gene	
AUTHORS	cluster has been deduced by comparison with similar gene clusters	
TITLE	in most cases and may be revised in future.	
JOURNAL	location/Qualifiers	
REFERENCE	1. 210614	
AUTHORS	/organism="Streptomyces rochei"	
TITLE	/mol_type="genomic DNA"	
JOURNAL	/strain="7434AN4"	
REFERENCE	/db_xref="taxon:1928"	
AUTHORS	/plasmid="pSLA2-L"	
TITLE	/note="linear plasmid"	
JOURNAL	1..1992	
REFERENCE	/note="left terminal inverted repeat, TIR-L; shows 99.4 %	
AUTHORS	(1981/1992) sequence identity to TIR-R (complement	
TITLE	(200623..210614) "	
JOURNAL	683..2188	
REFERENCE	/note="N-terminal sequence is almost identical (435/437)	
AUTHORS		
TITLE		
JOURNAL		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
REFERENCE		
AUTHORS		
TITLE		

with that of ORF143 at the right end of pSLA2-L until the inner end of TIR-L  
ORF1 (501 aa)  
similar to AL590463 Streptomyces coelicolor putative helicase, SCPI.136 (879 aa); homology is seen until the inner end of TIR-L"  
/codon\_start=1  
/transl\_table=1  
/product="putative helicase"  
/protein\_id="BAC76459.1"  
/db\_xref="GI:30698346"  
/translation="MSITSRTDQREAAQOEAVDVAVRALIEPVRLAPERGLTQIVINATGSKTRVARSAEKLRAGVLYVPSLDLTOTBAEMRAGTGPMIGSISGEISVAFPTDVEELINVMVPRPKVYVATYSLGDTLRBAHQGLPGMDLTIVDSAPHRSGRLKPAVAVHNDTRIPSLRLTMATIPRLMQDDEBEGAPGLVASMEDPGGLRARCFTLLISEA1IDRGI CAPYQVCVDITDQLQAQLGLVGESEVARGAIALAQTRLLKASSENFRRTLVFHMVKEASAFAAGLDVAKRLHAAQSGDYPTIWMNVCGEIRKPHRRRRLVIGFTSDITGTVVKEGFLGSKVYKVLGEQDTRCDSVYADVAGMPDGVQAVRRLRIPGGQKVASIIVVPLIDGEMADMLTSRPYNGALALPRQAPLLTGHGEEGRPAAPFPEFEVRPVLYGVGAQGSAGSAGSISITPSMSISIMRA"  
/complement(2315..3595)  
/note="ORF2 (426 aa)  
similar to AE004736-10 Pseudomonas aeruginosa hypothetical protein (442 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="putative secreted protein"  
/protein\_id="BAC76460.1"  
/db\_xref="GI:30698347"  
/translation="MRKRPRALGIPGLVAVLAPAPAAHSAOTGVYVTAAPPTAAPASTVEEDTLQVTLVAGSLRIRFRFNREGTSPLTIGTHAARAQGGATVADPGTDVVRFGGRSATLAPGTQRMSDPVALPTTAGGDVLYSLTLPQRTBSTYHSAAYOHNVAAAGDVTAAGDLPVSTATISWHFLSGSVSDRRKAGTADSSVTLGDSITDGEHTLTDVRRPVDVLAERLRDGLAGTVGNVAGIQGNRLRDEVPDPSGAASEFAVGEHSALEKQVATCTTPFAGDTIGYETRRERBAVRQOVNDWIRTSGAFTDVLFDVADLPAPADHLLPAIDGGDGLAPNDAGMAAARAAPPLSLR"  
/complement(3659..4645)  
/note="ORF3 (328 aa)  
similar to Y00459-2 Streptomyces griseus regulatory protein, Strr (350 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="Strr-like regulatory protein"  
/protein\_id="BAC76461.1"  
/db\_xref="GI:30698348"  
/translation="MTALAGNVLDQERTSIPISSLVPGYSPRLCGEDAEHVARLADTTEBALPIVSESTLRIYDGMHVRVLAAKAGHTTIVRLIEDGAESAFLIYASNNTHGQPLSRDRAAQRILIAQMPHPSDAVAVGIAIGKTYAALSLPAAQTPNQAQRGRGTRIRPLDGTGRKAELLLAORPQASVEVRAHAGISPATISVDRRLRAGRSRVEPVEWTPGARPGVGGSRATPGGGAPEVSVVRIIRPEBSPLVRLDRPLRKESGRIRLRLQCGAVERTALIAMAQVPPHCTDLVAELAREYADIMAEFAREVARTDG"  
/complement(4922..5812)  
/note="ORF4 (396 aa); lankacidin biosynthesis protein similar to AB004625-4 Pseudomonas aeruginosa pyrrolquinoline quinone biosynthesis protein B (304 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="pyrrolquinoline quinone biosynthesis protein B"  
/protein\_id="BAC76462.1"  
/db\_xref="GI:30698349"  
/translation="WILLGTAAQGGFPQWNCALCARGRGELPBAQSGECVAVSGDGLDMWILNAPSIRITQLLAAPALTPGPGRPDPVGVLLTDAEVLALGLATVRAQTGITYVAAPVAGLSAELPVRLGLDRYAPMDMDYAPGGFVAGAGLTVYAHVGTKAPYAHADPPAPVCAVRIEDPANGALVYAPCLATPMDGFDLLASATCALDGTFFSGSLGATITSAGAGSGLMKHLEPVAGRGSLALAHKRGRLRYTHLANTNPILIDBSAAHAAVAGAEVLPDSESLVLT"  
/complement(5815..6909)  
/note="ORF5 (364 aa); lankacidin biosynthesis protein similar to AJ27117-5 Gluconobacter oxydans putative pqq protein (359 aa)"  
/codon\_start=1  
/transl\_table=1



```

Db      2882  GCCCGAGACAGCCCGCGGAGCTGGTTCAGAGGCGGTGCGAGACGCGCTACACCGCGCGGGGGC 2941
QY      292   CGACCGTGTGTGTGCAACCGCGAGATGG-----CGCGCCCGATGAGCGCCG 339
Db      2942  CCGAGGGTGCCTGTGGTGAACGTTCGAGCGGACAGACACGACCGGGCGCGGTGCGGGCGT 3001
QY      340   CACTTCGATGCGGTGAGCGGACACCGCCCTGTTCATCTGTGTCTCTCTCTCTG----- 390
Db      3002  TGTCTACCGCGCGCGGTGTGCGGCGGGGAGACCGCGGTGTCTGTCTGTCTGTCTGTGCAACCGATC 3061
QY      391   -----CGCTCGCGGAGGGCGGTGTCTGTGACGACCCCGAGCTGAGACACCTCGCGTTGG 444
Db      3062  GTCCGCGCGACCCCGACCCCGCGCGGTGCTGTGTGCGCGCGCGCGAGCTTGTCTGTGA 3121
QY      445   TTCAGAGCGCTCGAGCGCAGCCGAGATTCAGACGTGCCCTGTGTGCTGTGTGACAGAGGACGCG 504
Db      3122  CCCAGGCGGTGTGCGCGACGCGCTGTGCGCGCGCGCCCGCTGTGTGTGTACACCGGGGGCGCG 3181
QY      505   CCGCGGTGACCGTTCGAGACGACGTCGATCCGCGCCGAGGCGCATGTGTGTGTGGGCTTCGCGC 564
Db      3182  TCTCCGTGTGAGCGCGCGGACGACCGCGACGAGCGCGCAGGCGCAGGCTGTGGGGGTTCCGCC 3241
QY      565   GGGTGTGTGGGCGTGGAGTCCCGCGCGCGCGGTGGGGTGG 600
Db      3242  GGGTGTGCGCGCTTCGACTGTGCGCCCGCGCATGTGTGGGGG 3277

```

QY 292 CGACCGCTGCTGTTGTTCACCGCGCACTCG-----CGCCCGGAACTCGCGCCG 339

Db 39233 CCGAGGAGCGCCTTGATGACCGCTTCAGCCGACCGACACGACCGAGCCGAGCCCGCGCTGCGCGCGT 39292

QY 340 CACTTCGACGCGCTTCGACCGGACACCGCGCTTCCACTGTGGTCTCTCGCTCG----- 390

Db 39293 TGTCTCACCGGCGCGGTGTGCGGCGGAGGCGACCGCGGCTGTGTGTCGTCTGTGCACTCCGATC 39352

QY 391 -----CGCTTCGCGGAGGCGGCGCTGTGCAACAACCCAGCGCTTGACACCTTCGCGTTG 444

Db 39353 GTGCGCCGACACCCCGACACACCGCGCGCTGTCTGTGCGCGCCCGCCCGACGTTGTCTGTGA 39412

QY 445 TCCAGGCGCTTCGCGCGACGCGCGGATTCGACGTCCTCCCTGTGCTGTGACACGAGACCGCGG 504

Db 39413 CCGAGGCGCTGTGCGCGACCGCGCTGCGCGCGCGCGCGCTGTGAGTGTTCACCCCGGCGCGCG 39472

QY 505 CCGCGCTGACCGCTTCGAGAGACGATGATTCGCGCCCAAGCCATGTCTGTGAGGCTTCGCGC 564

Db 39473 TCTTCGTCGAGCCCGCGCGAGACCGCGCGACGAGCGCCAGGATCTGTGAGGATTCGCGC 39532

QY 565 GGGGTGTGGGCGTGAAGTCCCGCCCGCGGAGGATG 600

Db 39533 GGGTTCGCGCGCTTCGAACTGCGCCCGCGCATGTGGGCG 39568

RESULT 12					
AX697977	AX697977	60196 bp	DNA	linear	PAT 02-APR-2003
LOCUS					
DEFINITION	Sequence 1 from Patent WO03010193.				
ACCESSION	AX697977				
VERSION	AX697977.1	GI:29499034			

RESULT	13
AF263912	
LOCUS	123580 bp DNA linear BCT 24-MAY-2000
DEFINITION	Scleractin nourselet ATCC 11455 myxobacterial gene cluster,
ACCESSION	complete sequence.
VERSION	AF263912 AF263912 GI:8050835

SOURCE ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
Micromonospora carbonacea	1	Farnet, C.M., Staffa, A. and Yang, X.	Genes and proteins for the biosynthesis of rosaramicin	Patent: WO 03010193-A 1 06-FEB-2003;	
Micromonospora carbonacea				Ecopia Biosciences Inc. (CA)	Location/Qualifiers
Micromonospora carbonacea					
Bacteria; Actinobacteridae; Actinomycetales					
Micromonosporinae; Micromonosporaceae; Micromonospora.					

```
source 1..60196
/organism="Micromonospora carbonacea"
/mol_type="unassigned DNA"
/db_xref="taxon:47853"
```

## ORIGIN

Query Match	20.7%	Score 124;	DB 6;	Length 60196;
Best Local Similarity	55.0%;	Pred. NO. 0.0012;		
Matches 350; Conservative	0;	Mismatches 250;	Indels 36;	Gaps 4;

**JOURNAL** Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim  
N-7489, Norway

**FEATURES** Location/Qualifiers

QY	1	CCGCGCTCCGAGAGCGTACACCCGCGGGGTGAGGTGCACT-----GGCGTACCGCAG	54
Db	38933	CCTCGCGCGCGCGAGAGCGTTCAACCCAGGGGGGTGCGCGGTGACCTGATGCCGAGGCTCTGGCCG	38992
QY	55	TGGGTGAGAGAGCGCCCGGTGCACTTGCCGGGTCTACCGGTTCGAACAGACAAACTTCTGGC	114
Db	38993	GGGCGCCGCGCGGTGCGCGGTGACCTTGCCGACGTACCGGTTCCAGCGGACAGCTACTGGC	39052
QY	115	TCGCGGTTCCTCTGGGCGCGGTCCCCGACA---CCGACGACGAGTGGCGTTACCAAGCTG	171
Db	39053	TGAGACCCGCGCGACAGCGGACCTCCGAGCCCGCGCGGAGACGCGTGGCGCTACCGGGTTCG	39112
QY	172	CCTGGCAACCCGCTGCACTCGAGCGGATCTCCCTGCGCGGACAGGGATCTGATGATGACCG	231
Db	39113	GCTGCGGCGGCGCTTCAGGCGCACCGGCGCGCGCGCGACCGGTGGCTGTGTGTACCG	39172
QY	232	GAGCGGACGTACCCCGCGCGCTTGACGGACGTGTCCGCGACGCGCTTGAAACACGCGCGG	291
Db	39173	GCCCGAGAACACCGCGGAGGCTGTGTGCAAGGCGGTGCGCGACGCGCTCACCGCGCGGGCG	39232

```

/organism="Streptococcus nourisae ATCC 11455"
/mol_type="genomic DNA"
/strain="ATCC 11455"
/db_xref="ATCC:11455"
/db_xref="taxon:316284"
complement (46.783)
/gene="nysF"
complement (46.783)
/gene="nysP"
/function="presumably post-translationally modifies ACP
domain on PKS"
/note="putative 4'-phosphopantetheine transferase"
/codon_start=1
/transl_table=1
/product="NysF"
/protein_id="AA071762.1"
/db_xref="GI:8050836"
/translation="MIEILPATVATEAYADRRPGDRLSSRBYIARAVESRQR
FTYHLARRRLRRLGHPDRAAILNRRKADPWPQGIAGSMTHCGAYDAAASPSLSA

```

gene  
CDS  
AVSIDAEPNGPPLPAGVLAIALPESRPHLVALAHRPDVHMDRLTFSAKSEYFKAWP  
LTQRELDPSEAEVIDPTQGAFTARLLVPGPLGGRVTVFPGRMHSTALLITVAHL  
PATPRDRREHETHLTVNSPLPRPTFG"  
complement(867. .2684)  
/gene="ny8c"  
complement(867. .2684)  
/function="presumably involved in efflux of nystatin"  
/note="putative transporter (ABC family)"  
/codon\_start=1  
/transl\_table=11  
/product="Nysg"  
/protein\_id="AAF71763.1"  
/db\_xref="GI:8050837"  
/translation="MSAPDLSEERTAPRPARIVGLLRPHRSVALAAMGVGIVL  
NAFGPLLRGVTLADGVLCGYPGAPGIDPAITRLLVLAIVYASLMLAQGR  
LVASAVRTIHEDRAREKTRPLPHDRKOPAGELSRRTNDIDNLOQTQITAE  
LITTS1ESLITMLVLMVISPISLAVMLSVPSALIAKISRAOPHVAOMFANGTL  
NAHBEVCTGHALIKGFDRAAABERFDACNDVYRAAKAQPASGAMPEVANGTL  
GVYLVAVIGAMVINGTITLGDVQAFILYAROSOPIVEIASVAGRLQSGIASQAVF  
TLDDAPEQADPLRPGTPARAGRVETDVSFSPDPLINLSLTVRPGSTVAIVG  
PTGAKTTIGNLLMRPEPDSGRILLDDTDTMTTRDILSRRLGVLQDTMLFGGTTA  
ENIATGAPGACRADIIEAARATCADRFITLQGYDTVLDDSGTVSABKQLTVAR  
AFLARPVAVLDEATSSVDTRTEVLIQRAMSLRAGRTSPVIAHRLSTIRADLLVM  
DAGRIVEQCTHQLCAQGLYARLHAARTHTPTAGAAG"  
complement(2662. .4416)  
/gene="ny8h"  
complement(2662. .4416)  
/function="presumably involved in efflux of nystatin"  
/note="putative transporter (ABC family)"  
/codon\_start=1  
/transl\_table=11  
/product="NysH"  
/protein\_id="AAF71764.1"  
/db\_xref="GI:8050838"  
/translation="MLLRALRQALRPYAWATAALVALQVLIGTLLPTLGAALIDQ  
GVRRGGRITELGVGVVAVVQIAALAAALAAATATMGRDRLSALPRITLDES  
AEIRGCRPSLITRSVNDVOVONLAAQPGFIVCAPIKMGSGITLALRQVPLALL  
LVALVYVAVCGCLLAKNGITYARQTLTDLRLGRLREAITGVRVRSFVDDHERA  
RFAQTNDAFLVSRVGRLLIATMLPVVLLANGFVALLMTQSHRIDARMPIGISIA  
LISYLSLIMSVMVAFLSVPRVACGRIAEVLDTGSSVAPPAQPVPGPAGRI  
ELCAAGYRPGABEPLRDVLTVEGERIEAVTSGSGKTLINLVLELATAGVAV  
RYGTDVRELTAATLAAVGFYRORPYLPGSTVANLRCRGPADTDERELARVQA  
ADPVARMPGDLAEITQGGANVSGGQRRLSLARLLRPRITLDDCCSALDQATDA  
ALRTALVPTTACATVITVAQRISAGDADRIVLDRGVVAGCTHVDLRTSPYREI  
ALSQTEBAAGTLAGRP"  
4714. .5748  
/gene="ny8D111"  
4714. .5748  
/gene="ny8D111"  
/function="presumably involved in mycosamine biosynthesis"  
/note="putative dcgp-mannose-4,6-dehydratase"  
/codon\_start=1  
/transl\_table=11  
/product="ny8D111"  
/protein\_id="AAF71765.1"  
/db\_xref="GI:8050839"  
/translation="MSKRLITGITGQDSYLAELHLSQGYOVWGLIRGOANPRKSRV  
SLIASLELPIDGLMDQSLVAVDTVOGDEVYNGAISFVPMWQOALVTEVANGV  
VLRMLEAIRMVGLDSTSRVSPRGQIRFYQASSSEMFGAETPOREBTLFPHRSPYG  
AKAAGYITTRVRSFGWVAVSGMLFNHRSRPGORPRTKISLAVARIKQLODKL  
ALGNLDAYRDKGACDYVRAAHMLMCOQDAGDVTYGTQGMVSRVAVRTAFHEVGLN  
EDYVIDPDLVRPAEVEVLCDSATAQDLKPKVDVPTLKRMMVDSDLQAVRENO  
YGDVLLAANM"  
5930. .34363  
/gene="ny8I"  
5930. .34363  
/gene="ny8I"  
/function="responsible for condensation steps 9 to 14 in  
the nystatin polyketide backbone synthesis"  
/note="polyketide synthase"  
/codon\_start=1

/transl\_table=11  
/product="NysI"  
/protein\_id="AAF71766.1"  
/db\_xref="GI:8050840"  
/translation="MDNEQKLDYUAKLTADLRTRRRVHKHLSAAQEPVALIGMTCR  
YFGVRSPEPDLMRWYAGEHGVTPPTRGOMLELMAAPTSQGGTFLHAPDPDAPF  
GISPREAVAMDQQRVLSAEAEERAGIDPTSVSGSTGVGIGAMQDVRGADP  
AGFPOLTNGTGSVLGRLSYTFGTGVPATVDTQSSSLVAHIALTOALRAGCTIAL  
AGYITMGRPFRIEMRGQGLSADGRCSREDPTADGTCMAGCVILVLERSDAVRN  
GHEITLAVNRGTAVNDGASNGITAFNSQOVITQOALNARLACADIDIVELHGTG  
TLGDVEAQLATATGQNRPADRPLLSVKSNIHTQAAAGVAVIKVMAMRHGTL  
PRTTHAEPTHHVDVSGCAVRLITDTTWPAGSRRAVASFSGISNAHTIIEQAP  
EPOEDAAATQDDAAGSTPATPVPVGVVPLPAGTRPALGSOAAALAPADTGRP  
DILDLASLATRAGFEHRAVILATDHALTDGLTALADDDPAAPADITCTTAT  
RLAVIPTQGAQRILACGELARFPARTDAAALDAPRPHLDREBLWKTDAALL  
DRTAYQALPFAVEVALYRLIESFGVRDHLGSHVGRIVAAHLGVSLLAATATVA  
AAGRLMOALPDGAMIAVQASADVAFLAGHEDVAIAVNGPSAVVLSGAEATVA  
LAEOALACQKTRRLRVSHAFSPIMEPMLDFAVVEEDTLQDPLPVSNITGKPA  
TVAQTSADYVWDRHVAVRPADGIDMLABHDTTAFLELPGDGVISAMQDCLDADA  
DAVTLBALRAGREHRTLTTLALGHVNGATLDMGCPAGTCARPTDPTVAFORRY  
MKALQSTADLRSGELAAHHPILSAAVSLADAGCTLLTGRLSKOTHPMLADHTYR  
TTLPGTAFLELAVAGBVGCDRVEBELPAVPLLPQGGVQVQVLMIGNPDVSGRT  
VVHARPTGDTPTWTAATGVLTTADASRQLPASBOGTEPLADPHPLADQAMP  
AGABELPLDGHYRLADGFGYGVPGFGLRAAMRGDVAIVAEVLEPAGRSAPAEGL  
HPALDALLHAAPFTGLGRRGGLPFSMEGSLHAGACTTLRVLTPVADALALT  
ADGTAPLYLSDSLVLSVATQOOLTTAAVANDLPRLDMPVQPTAPDRCGVALLGA  
DPGGLTTAGPADAPAYPDLALAAADPEPTTYSLTAGTDDAADPARSKARCAAE  
ALAAQVTLDDHDERFAARLVFVTRGATVGRVAAAASWGLVRSQSENPQCFALVDL  
DPSGEGAVLVTGGTGGCAVLAHLVAEYGRVLDLVSRSGERAVAGELVAELAGV  
GARVAVCAVDLDRAAVVELGVGHVSAVVHAGVLDGDMVALGERSLAVLRPVV  
AVMHLHEATRGDILDAPIVPSLAGVPSRGQANAAAPRLDMPVQPTAPDRCGVALLGA  
LAKPWSGSGMTGTLTVYDABRLARSGVPLSAQGLALPDAVAGTDATCVPVRLD  
LPLVLRABEVEPLRLSLIRVARRAASVAGSAPAGLALRLRLDGBDEMYLDVRG  
OVALVLAGATGADVAGARFRLGDFSLTAVALRLRNLVTGRLPATLVEFYPYRTH  
LATVYDELLGTDAEVAIVQPAVANADDPITYVMACRPGVGSPELMLVTEGT  
DAVSGEPNRRGMDVSLYHPDHPGTSYTRSGELHAGREDDPFGPMSPREALATD  
SQOQLLESMEALIRADITDVSLSGSTRGVPAGTMSDYDSMLASPEREGQSGSS  
PLSAGRVATLGLGPAVYTDVACSSSLVAMHMMQALRSSEGLALAGVTVMSGP  
AVPFAAQRIISPGRKAPADADAGVSGVGLVLEBOSDVRNGHEITLAVNGP  
SAVNDGASNGITAFNSQOVIRLOALSGGLTADGVDAEAGHGTGTPRIBQA  
LLATYGRREPERPLPLGVSXSNLGHTOAADGAVAGVAKMVLAVAGTVPRLHVAO  
SHVMSBEAVLLESOAMPERTGRRVRSVRSFGISGNVAVIYQAPAKRTIAAGA  
ARTPGAVPVLISGGRSALGQARLLIGQARDAELVYVALSLATTRRSFEQQA  
VVAQDRDILASLALADRDPDAVBEBAERGAVALPFGQSSQRAAMGSELHQV  
PEFAAPFAPVACAVDPDLRLPREVPEFIDESDEBALDETGTOPALPFAVVALFRL  
VESWGVREDPVAAGHSIGELAAHVAGVLTLEDACGLVAAARTLMOALPGCAMIAIOA  
TEDELAHLDDTVAILAANGPOSVITSGDEBAERTIATPAREGKTRIRVSHAFS  
PRMGMDLAHPRIVAGEGLTYRARPITPVSLDGRRDDAVCTAETWVGHVRAVAFD  
CVRTLRDAGATTLLELSDGLITMAEDTLGDBDHAEVPMIRAGABEHLAATLAR  
LQVRGVDDMAVYLAGTGARTDLPYVAFQHAAYWPOLEPTPAALAAADPADQOQLMA  
VERGABRELDIILIGEBQDLTPDLISLALTSWRGNOBEKHLIDTLRAREVETRLSKP  
TAPVLDGMLVVASADPADOPALLDGLAADGAGVRLRLDSDCAGRVILARL  
ARTADVAAATQVLSVLPDREDDDDCPDLTGLALITVALYQALLADTQAGRLVATRG  
AVSTRPADPTHPQAAAKWCGRGYALAEHPRLKGLVYLPQVYPERBAQRAGITIAVK  
DAPDEDOVALAKATGVSGRULVRITVEALPFAEETATGTVLTGCTGALAEARWML  
ARAGAQHLVLTSRGPPAGAAERLAELEGGPSVVAACVABDALAAVYALPABE  
LPLTGVHTAGVGHYGLDPTLSTAEFACITLAKKAGANMLALADRELDPEVLEFSI  
LAGVSGNGQSYGAAANAYLDLALHRRARGLAATSVAAGPAKAMADMDVASETLR  
QGLDGLDAPAMTELIRAVVRODVTYVADVOMYALPFSARSAIILAGIPEVAL  
AADRTBEDAPASBEVYRVRALABEQLRLITDLPVRESVATYCHSADLVBERAR  
RDVGDSDITAVELIKRQALATGSLSPSTWVDYTPPLELAQYLRAEITAGVLEVAGV  
ATGADSDPILAIIGMAKPEFGVSSPEQMLVASGDTALISFPVNRWQGHLEDPD  
PDRPGYITSTYSGFLHEADPEDPFPGISPEBALVMDQOQLLETWESPERAGIRP  
ETLSTLTGTFTGVSQYSEYGAAGDGEHGHVTVSSPSVLGRLSYVGLGSPATVD  
TACSSLVAHLACOSLANSGBNLVAGGATIMTTPNPFIAFSORLALAGRCARFAS  
DDAGMTLABEGVAVILERSLADQBNHGPAVAVRGAISNGLTJANGSCOR  
VTRQALNARLAPRDIIDLBHAGCTPGDPIRQOALPATYGRDRDPESALLASVKS  
NIGHTQSAAGTASYIKVMALRHSLEPPTHADSPSHVDSAGVRLTALLARABET  
GRPRRAVSSFGISTANAVILEQAPVADTAEBERPAVAPVIAAGVAPWVYASAA  
ALRQCAERLLAHARTVGALEPAQPLDGLSLVBARAFEBRAVAVVPGAGDPLALR  
AVADGSPVAVRAGVADGERTVVFPGQSGOMVGMGQLDDESVPFRABCAAL  
ABFTDMSLVDTLRGVYAPSLERVDVQPAFVAVVSLAALMRSGVLPDAVVGHSQG

[illegible]

JOURNAL Patent: WO 03010193-A 15 06-FEB-2003;  
 FEATURES Ecopia Biosciences Inc. (CA)  
 Location/Qualifiers  
 source 1..11238  
 /organism="Micromonospora carbonacea"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:47853"

ORIGIN

Query Match 16.7%; Score 100.2; DB 6; Length 11238;  
 Best Local Similarity 52.9%; Pred. No. 0.33;  
 Matches 250; Conservative 0; Mismatches 208; Indels 15; Gaps 1;

QY 143 CACCGGACGAGTGGGCTTACCACTGCGCTGGCAACCGGTGACCTTGAGCGGTCTC 202  
 DB 2949 CACTTGAAACCGTTGCGACCGACCCCGACCCCTCAACCAACCGCTGCTGAT 3008  
 QY 203 CCTGGCGGACGGGCTCTGTGTGACCGGAGCGGCACTACCCCGGCTGGAAGAGT 262  
 DB 3009 GGTGCCGAGGCGCTGGCGGTGACGGGGTGTGGCGGGGTACGGAGGCGCTGGCGC 3068  
 QY 263 GGTCCGACGCGCTGGAACAGCGCGGGCGACCGTGTGTGACCGCGAGTCCG 322  
 DB 3069 GCGTGGGCTCTCGTCAACTGCTGACCGTCCGACCGCGACCGGGCGGCTTGGCGC 3128  
 QY 323 CGCCCGGATGGGGCGGCACTGACCGCGGTGACGGCAACCGCTGTCACTGTGTCTC 382  
 DB 3129 GCTCTGACTTCGCGGCGCCCGGACCGGAGCGGCGCGGCGGCGGCGGTGTCTC 3188  
 QY 383 TCTGCTCGCGCTGCGGAGGCGGTGTGTGACGACCCGACCTG-----428  
 DB 3189 CCGTGTGGCGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 428  
 QY 429 -GACACCTTGGCGGTGTCAGGCGGCTGGGCGGAGCGGAGTGAAGTCCCTGTGCT 487  
 DB 3249 CGCTGCTGCGCTGATCAAGGCAATTGGGAGGAGCGGAGGAGCGGAGGCGGCTTGGGC 3308  
 QY 488 GGTGACCAAGGAGCGCGCGCGCGGTGACCGTGGAGAGAGAGTGTGCGGCGGAGGAT 547  
 DB 3309 GGTACCGGAGGAGCGGCGGCGGTGTCTTCCGAGAGTGCAGGAGCGGCGGAGGCGCA 3368  
 QY 548 GGTGGTGGGCTGCGCGGAGTGGGCGGTGAGTCCCGCGCGGAGTGGG 600  
 DB 3369 GGTGTGGGCGCTGCGGCGGCTCGGCGGCTCGAATGCGGAGCGATGGGCG 3421

Search completed: April 7, 2006, 16:57:34  
 Job time : 2906.98 secs



**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd

OM nucleic - nucleic search, using sw model

Run on: April 7, 2006, 02:24:14 ; Search time 2904.81 Seconds (without comments)

(Without alignments)  
11760.789 Million cell updates/sec

```
Title: US-10-611-442-2_COPY_1200_1800
Perfect score: 601
Sequence: 1 gcctcctacgcggcgga.....ggacgcctcaccgcgcact 601
```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ,

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

```

1:  gb_ba: *
2:  gb_in: *
3:  gb_env: *
4:  gb_om: *
5:  gb_ov: *
6:  gb_pat: *
7:  gb_ph: *
8:  gb_pr: *
9:  gb_ro: *
10: gb_sts: *
11: gb_sy: *
12: gb_un: *
13: gb_vi: *
14: gb_htg: *
15: gb_pl: *

```

**Pred.** No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	601	100.0	47981	1	AF263245	AF263245 Micromon
2	601	100.0	47981	6	AX112026	AX112026 Sequence
3	397.8	66.2	20235	1	SEREXYAB	M63677 S. erythrae
4	397.8	66.2	20235	6	AR049368	AR049368 Sequence
5	397.8	66.2	20235	6	AR095529	AR095529 Sequence
6	397.8	66.2	32299	1	AY661566	AY661566 Saccharo
7	378.4	63.0	9769	1	SERERA	X56107 S. erythrae
8	378.4	63.0	20444	1	SERERAB	X62569 S. erythrae
9	281	46.8	31531	11	AY771999	AY771999 Synthet
10	263	43.8	61845	1	AY628658	AY628658 Aetomicro
11	243.6	40.5	210614	1	AB088824	AB088824 Streptomy
12	220.2	36.6	4725	6	AX697993	AX697993 Sequence
13	220.2	36.6	60196	6	AX697977	AX697977 Sequence
14	210	34.9	104326	1	AB070940	AB070940 Streptomy
15	210	34.9	110000	1	BA000030	Contribution (37 c
16	209.8	34.9	58343	1	AB089954	AB089954 Micromon
17	207.6	34.5	9684	6	C0875927	C0875927 Sequence
18	207.6	34.5	167179	1	AY899214	AY899214 Streptomy

19	207.6	34.5	164051	6	C0875911	Sequence
20	205.8	34.2	68451	1	AY330485	Saccharop
21	204.4	34.0	5355	6	AX697995	Sequence
22	202.6	33.7	11219	1	SEREXYAA	M65676
23	202.6	33.7	11219	6	AR049367	Saccharopol
24	202.6	33.7	11219	6	AR095528	Sequence
25	201	33.4	76196	1	AY545415	Sequence
26	201	33.4	84985	1	SNA278573	Streptomy
27	199	33.1	24444	6	C0875915	Streptomy
28	198.8	33.1	11393	1	AF357202	Sequence
29	198.8	33.1	11393	6	AX703543	Streptomy
30	198.4	33.0	78210	1	AB070949	Sequence
31	198.4	33.0	110000	1	BA000030_05	Streptomy
32	197.2	32.8	110000	1	BA000030_04	Continuation (5 of
33	195.6	32.5	9513	6	AX089460	Continuation (5 of
34	195.6	32.5	50000	6	AX089417	Sequence
35	195.6	32.5	50000	6	AX089420	Sequence
36	195.6	32.5	80161	1	AY007564	Saccharop
37	195.6	32.5	80161	6	AR165018	Sequence
38	195.6	32.5	80161	6	BD137649	Sequence
39	195.6	32.5	80161	6	AR281866	Bioynthetic
40	190.6	31.7	74787	1	SPAS80915	Sequence
41	190.6	31.7	74787	6	C0839963	Streptomy
42	186	30.9	75236	6	AX600586	Sequence
43	185.2	30.8	138203	1	AY10323	Sequence
44	184.2	30.6	27541	6	AX117106	Sequence
45	184.2	30.6	123580	1	AF263912	Streptomy

## ALIGNMENTS

```

RESULT 1
AF263245
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

AF263245 47981 bp DNA linear BCT 17-SEP-2000
AF263245
gene cluster, partial sequence.
AF263245
AF263245.1 GI:10179840

Micromonospora megalomicea subsp. nigra
Micromonospora megalomicea subsp. nigra
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporinae; Micromonosporaceae; Micromonospora.
1 (bases 1 to 47981)
Volchegursky, Y., Hu, Z., Katz, L. and McDaniel, R.
Biosynthesis of the anti-parasitic agent megalomicin:
transformation of erythromycin to megalomicin in Saccharopolyspora
erythraea
Mol. Microbiol. 37 (4), 752-762 (2000)
10972798
2 (bases 1 to 47981)
McDaniel, R. and Volchegursky, Y.
Direct Submission
Submitted (03-MAY-2000) Kosan Biosciences, Inc., 3828 Bay Center
Place, Hayward, CA 94545, USA
Location/Qualifiers
1..47981
/organism="Micromonospora megalomicea subsp. nigra"
/mol_type="genomic DNA"
/strain="NRRL3275"
/sub_species="nigra"
/db_xref="taxon:136926"
complement (1.. 144)
/gene="megT"
complement (<1.. 144)
/gene="megT"
/codon_start=1
/transferable=1
/product="TTP-4-kefo-6-deoxyglucose-2,3-dehydratase"
/protein_id="AAG1396.1"
/db_xref="GI:10179841"
/translation="MGDRVNGHATPESGAIKFLTRHGCPPTATDDVHMLAHLAAAE

```

gene  
CDS

928..2061  
/gene="megDVI"  
928..2061  
/gene="megDVI"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
/protein\_id="AAG13907.1"  
/db\_xref="GI:10179842"

gene  
CDS

2072..3382  
/gene="megDI"  
2072..3382  
/gene="megDI"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-megotamine glycosyltransferase"  
/protein\_id="AAG13908.1"  
/db\_xref="GI:10179843"

gene  
CDS

3462..4634  
/gene="megY"  
3462..4634  
/gene="megY"  
/codon\_start=1  
/transl\_table=11  
/product="mycarose O-acetyltransferase"  
/protein\_id="AAG13909.1"  
/db\_xref="GI:10179844"

gene  
CDS

4651..5775  
/gene="megDII"  
4651..5775  
/gene="megDII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-3-keto-6-deoxyhexose 3-amino transferase"  
/protein\_id="AAG13910.1"  
/db\_xref="GI:10179845"

gene  
CDS

5822..6595  
/gene="megDIII"  
5822..6595  
/gene="megDIII"  
/codon\_start=1  
/transl\_table=11  
/product="daunosaminyl-N,N-dimethyltransferase"

928..2061  
/gene="megDVI"  
928..2061  
/gene="megDVI"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
/protein\_id="AAG13907.1"  
/db\_xref="GI:10179842"

2072..3382  
/gene="megDI"  
2072..3382  
/gene="megDI"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-megotamine glycosyltransferase"  
/protein\_id="AAG13908.1"  
/db\_xref="GI:10179843"

3462..4634  
/gene="megY"  
3462..4634  
/gene="megY"  
/codon\_start=1  
/transl\_table=11  
/product="mycarose O-acetyltransferase"  
/protein\_id="AAG13909.1"  
/db\_xref="GI:10179844"

4651..5775  
/gene="megDII"  
4651..5775  
/gene="megDII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-3-keto-6-deoxyhexose 3-amino transferase"  
/protein\_id="AAG13910.1"  
/db\_xref="GI:10179845"

5822..6595  
/gene="megDIII"  
5822..6595  
/gene="megDIII"  
/codon\_start=1  
/transl\_table=11  
/product="daunosaminyl-N,N-dimethyltransferase"

[illegible]

```
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AG13916.1"
/db_xref="GI:10179851"
/translation="MTRHVTLLGVSGVSGALREFTHPLRIARAVRCSPPQPS
AGIEHRVDLBERGVAQVADTDVYVHLYAVAGSTRSAATVBAEKNAGINED
LVVALRREPAPVLLFASITQANPAPRVAOHKI EARRILRQATEDDVDVILR
LPAIYHSGSPSGQTRGVVYAMIRALAGAPITMHGVSRYRLLHVEDVATFAL
HNHBAIVGVDTWSPSADBARPLGEI FETVAASVARTGNPAVPVSVPPENAEANDFR
SDPDFSTPEPTLGMHPRVPLAIGDIRTVAALSTKE"
12181..22821
/ gene="megAI"
/ cds
12181..22821
/ gene="megAI"
/ note="polyketide synthase"
/ codon_start=1

Query Match 100.0%; Score 601; DB 1; Length 47981;
Best Local Similarity 100.0%; Pred. No. 6,5e-52;
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCTCCAGCGCGCGGCAAGCGCTTCGAGAGCCCTTGCCTGCTGGTCCGCGCAGT 60
DB 37506 GCCTCCAGCGCGGCGCAAGCGCTTCGAGAGCCCTTGCCTGCTGGTCCGCGCAGT 37565
QY 61 GGGCTGCGCGGTCACTTCGATCGCTGCGGTCTGTGGGCGGGGCAACATGCGCGTACC 120
DB 37566 GGGCTGCGCGGTCACTTCGATCGCTGCGGTCTGTGGGCGGGGCAACATGCGCGTACC 37625
QY 121 GAGGCGCGGCACTAAGCTGCGGACGAGGCGCTGCGGCGCAATGACCCGACGCGGCGATC 180
DB 37626 GAGGCGCGGCACTAAGCTGCGGACGAGGCGCTGCGGCGCAATGACCCGACGCGGCGATC 37685
QY 181 GAGGAGCTGCGGCAACCCCTGGAAGCGCGGGGACCCCTGGGTGTGCGGTGTGACCTGGAC 240
DB 37686 GAGGAGCTGCGGCAACCCCTGGAAGCGCGGGGACCCCTGGGTGTGCGGTGTGACCTGGAC 37745
QY 241 CGGAGCGGTTCTGTGAAGTGTTCACCGCGCGCGCGCGCGCCCTCTTGAAGAACTC 300
DB 37746 CGGAGCGGTTCTGTGAAGTGTTCACCGCGCGCGCGCGCGCCCTCTTGAAGAACTC 37805
QY 301 GGTGGGAGTCCGCGCGCGGCGGAGAGACCGGATCGAGAAATCGATCTCGCCGCGGCTG 360
DB 37806 GGTGGGAGTCCGCGCGCGGCGGAGAGACCGGATCGAGAAATCGATCTCGCCGCGGCTG 37865
QY 361 GCGTCATATCGCGGAGCGCAAGCTCAAGACATGTGCGCGGTGTGCGGACCGCAAGTG 420
DB 37866 GCGTCATATCGCGGAGCGCAAGCTCAAGACATGTGCGCGGTGTGCGGACCGCAAGTG 37925
QY 421 GCAGCGGTGCTGGGCGCAAGCGGACGCGGATCGAGGCGTGAACGCTTCCGTTGAC 480
DB 37926 GCAGCGGTGCTGGGCGCAAGCGGACGCGGATCGAGGCGTGAACGCTTCCGTTGAC 37985
QY 481 CTGGGAATTCGACTCAAGACCGGCGGTGACCTTGCAGAAACCGGCTCGCGGCGTGAACCGG 540
DB 37986 CTGGGAATTCGACTCAAGACCGGCGGTGACCTTGCAGAAACCGGCTCGCGGCGTGAACCGG 38045
QY 541 GTTCGGGAGCGGCAAGACATGCTTGTGACCAACCGGACAGTGAACCGGCTCAACCGGCGAC 600
DB 38046 GTTCGGGAGCGGCAAGACATGCTTGTGACCAACCGGACAGTGAACCGGCTCAACCGGCGAC 38105
QY 601 T 601
DB 38106 T 38106
```

```
RESULT 2
AX112026 47981 bp DNA linear PAT 01-MAY-2001
LOCUS Sequence 1 from Patent W00127284.
DEFINITION AX112026
ACCESSION AX112026
VERSION AX112026.1 GI:13938925
KEYWORDS
SOURCE Micromonospora megalomicea
```

```
ORGANISM Micromonospora megalomicea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micromonosporinae; Micromonosporaceae; Micromonospora.

REFERENCE 1
AUTHORS McDaniel, R. and Volchegursky, Y.
TITLE Recombinant megalomicin biosynthetic genes and uses thereof
JOURNAL Patent: WO 0127284-A 1 19-APR-2001;
Kosan Biosciences, Inc. (US)

FEATURES
source
1..47981
/organism="Micromonospora megalomicea"
/mol_type="unassigned DNA"
/db_xref="taxon:47868"
complement(<1..144)
/ note="unnamed protein product; megBVI (megT),
TDP-4-keto-6-deoxyglucose-2,3-dehydratase; SEQ ID NO: 2=
translated amino acid sequence"
/ codon_start=1
/ trans1_table=11
/ protein_id="CAC37805.1"
/ db_xref="GI:13938926"
/ translation="MGDRVGHATPSTQSAIRFLTRHGGPTATDVIDHMLARAAE
HLE"
928..2061
/ note="unnamed protein product; megDVI,
TDP-4-keto-6-deoxyglucose 3,4-isomerase",
TDP-4-keto-6-deoxyhexose 3,4-isomerase; SEQ ID NO: 3=
translated amino acid sequence"
/ codon_start=1
/ trans1_table=11
/ protein_id="CAC37806.1"
/ db_xref="GI:13938927"
/ translation="MAVEBRRLRGRLQARGLYMGFGANGDIYSMLSGRDDPMTW
YERLRAGGRTAGTAVGDRHRAVLADPGTHGPDPARWQVACPAASPA
GPFREFYATTEPAASVTVADWLQQRCAVLTELSRFDLVDPARVPVLAAGTAPA
LKGVDPRRLRSWTSATRVCLDAQVSPQALVTEOALTALDEIDAVTGGDAVLGVV
AELAAVTGNAVLATVTELPALARLADDETRVTVTEVSTRSPGVHLERRTAASDR
VGVDPVTCGEVTVVVAANRDREVTDDPRDPRVDGSGAATLSSRSGSRITLDLALY
ATLATAALAAAAPVLPRLSRSGPVIRRRSPARGLSRCPVEL"
2072..3382
/ note="unnamed protein product; megDI, rhodosaminyl
transferase (eryCIII homolog), TDP-4-ketoglucose
glycosyltransferase; SEQ ID NO: 4= translated amino acid
sequence"
/ codon_start=1
/ trans1_table=11
/ protein_id="CAC37807.1"
/ db_xref="GI:13938928"
/ translation="MRVSPSMAVNSHLFGVLPIASAPQAGHREVVASPAITDVT
FGALTPVYGDVLEVMHAHAGODIVERMTLDWDQSHHTMSWDILGMOTTPPT
PRLASPDSLIDGMVDFCSMRPDMVWEPVLPAPITARVGTTPARMLMGCVATPR
APQSPFLRLAHQEVEREDPLAEWPDWTLRRRGDDPHLSFDEBLVIGQWVDPIPEPL
RIDTGRTVGRVYVNGPSVVPAMVLRPEERRVYCLTIGSGSRHEIGICVSGEMLD
ATADIDAEFVATFDQOLVGVSPANVTAFAPVPMVNLPTCAATVHNGSGSTMLTA
ATHGVPOILLSDADREVAHAKOIODTGAGISLVAGMTAPHLAIGERIVLDEPRVIRGA
EMRDEGMRDPSPAQVVGICQDLADRARAGQPRRTAEPHLPR"
3462..4634
/ note="unnamed protein product; megG (megY), mycarosyl
acyltransferase, mycarose O-acyltransferase; SEQ ID NO: 5=
translated amino acid sequence"
/ codon_start=1
/ trans1_table=11
/ protein_id="CAC37808.1"
/ db_xref="GI:13938929"
/ translation="MVTSTNLDLTARPALNSLTGMRFVAAPLVFFTHVLSRLIPNSYV
YADGLDAFWQTTGRVGSFFPLTSGFVLWTSARASDSVSPRRRVCCKLPNNLVAF
AAVVLFLVTVGAVSGBALIPNLLIHAMPALTEISGIVSPMSLACGAPFLCPFLP
LWISGIRPERLMMAVAAPVAPVAPVADLLSSPPLIGLSGLAEIADOMETVLP
ATRSLEPILGITLAILITLGRWINGLLPAVLLFPVFPASLPLEGVATISSMMLP
LVLLIASGATADLQGRITPRNRVWVWLDVSPALVWVFLVTVGADLLGFSQEDA
PLGLAFMLIPPLAIVSLVLSMLYRFVLELPVWRNARPASARKKATBEQTPSR"
4651..5775
/ note="unnamed protein product; megDII, deoxy sugar
```

transaminase (eryCI, Dnrf homolog),  
TDP-3-keto-6-deoxyhexose 3-aminotransaminase; SEQ ID NO:  
6= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37809.1"  
/db\_xref="GI:13938930"  
/translation="MTYVMSYLLERERRADILDAVQKYPAGSLLIQSVENFETE  
YARHGIACVGVNGNNAVKALAEVSGRDEVTVSNTAAPTVLAIDIGARPVF  
VYVDRDYLMDTDLVRAAVTPRTKATVPVILVGGCVMTALRELRGKIKYEDCQ  
AGARRDRLAGTMSDAASFPTKYLKGYCGAGAVTDDSTARKLRKRYGHEE  
VYVYTRGVNSRLDEVQASILRRKLTDLDAVYAGRAVAKRYVDGLADQSDHSL  
PVYVDGNEHVYVYVNHPRRDIKRLRDGYDLSIMSYPMVHTMTGFAHGVASG  
SLPYTERLAGEISLPMYPSLPHDLQRIEVAEVEVITGL"  
5822..6595  
/note="unnamed protein product; megDII,  
daunoseaminyl-N,N-dimethyltransferase (eryCVI homolog); SEQ  
ID NO: 7= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37810.1"  
/db\_xref="GI:13938931"  
/translation="MENSHTSTSTVAPYPERADITYDFTYHGKGYRABADLVYA  
RKTPQATLLDVAAGTGSHLVELADSFREVGVGDAAMALATARDPGRBLQGM  
RDFSIDRRFPVYVCMFSGTGLYDLELDRAVANLAGHLAPEGTLVYEPWFETFRP  
GMVAGDLVTSQDRRISMSHTVPAGLPDRTPASRMTIYTVGSPKAGIEHFEVYMTL  
FARAAYQAFORAGLSCSYVGHDLFSPGLFVGYAABRGR"  
6592..7197  
/note="unnamed protein product; megDIV,  
TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryVII, dnmU  
homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase; SEQ ID  
NO: 8= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37811.1"  
/db\_xref="GI:13938932"  
/translation="MRYBELIGCVFTFTQTFADERGVGTAYQEDPVFAALGRPLF  
PVAOVSTRRRGVGVHFTTPMGSMNAKYVCARGRAMPAVDIRGSPGTAPEV  
ELSAEWSVGLYLPVGMGHLFVSLDDTTLVYAGVYDPERAVHRLDELALPIPA  
DLIDVMSERDRAVTLLEARDQGLIPYACRAAAHVVRT"  
7220..8206  
/note="unnamed protein product; megDV, TDP-hexose  
4-ketoreductase (eryBIV, dnmU homolog),  
TDP-4-keto-6-deoxyhexose 4-ketoreductase; SEQ ID NO: 9=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37812.1"  
/db\_xref="GI:13938933"  
/translation="MVVLGASGFGSANTHALADLPYRVLVAREVVVPSGAVADYE  
THRYDLTEPGALAEVADARVAFPPAQRIGTSGMRISSEDDVARETNVGLRLIIV  
LSRSPHAPVYVPPGSPNQGVRTAGRVLDGSKQHPGAVDROGHTSOQLKEKTAAG  
AIRATSLALPPVFPVPAAGTADNRGVSTMLRRLLTQGPLTMHMDGTVRRLIYTTDA  
AIRAVTLADHADALAGRHLGTRSMPLGTFQAVSRSVARHSGEDPVVPSVPPA  
HMDPSDLRSVDEPARFTYATGMRAVTYMAEADVRYAALAPRAAAPSBS"  
complement(8228..9220)  
/note="unnamed protein product; megBII-1(megDVII),  
TDP-4-keto-L-6-deoxy-hexose 2,3-reductase; SEQ ID NO: 10=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37813.1"  
/db\_xref="GI:13938934"  
/translation="MGTTGASARVVRGSAALHTSRLMTGVNFSGRVTDGDAALMD  
HALRGVACIDTADIVGRLTKHTELVGMFCAQGRRETEVLATKESMERV  
DGLSARITVAACENSLRLGVHDHIDYQTHIDRAAPMDVMAAELVSSGVGVY  
GSSNLAGWHTAAQESABRNILGMSHCCLYLAVRPELDVYPAQAYGVFANS  
PLHGLSGVLEKLAGTAVYAGCAQRAOVLLPAVRPLVEAEDVCRRLGAPVAGLA  
MYLSRPGILGVIVIPRPEOLDALRAAELTGBEELAEALIPAPAVDVPV"  
complement(9226..10479)  
/note="unnamed protein product; megBV, mycarosyl  
transferase, mycarose glycosyltransferase; SEQ ID NO: 11=  
translated amino acid sequence"

/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37814.1"  
/db\_xref="GI:13938935"  
/translation="MRVILTSFAHRTFHQGLVPLAALHTAGHDVRAVQPELTDVYV  
GAGLISVPUSDRRLDTSPEAAQVHRITTDLPARRGELASWEVTLHISERTSEF  
VFPEVNNDSFVDELVEFADWRPDVLVMEFPTAGVAAACAAHARLHLSGLDLY  
FRSRGDLRQGRPADRPDLGMLTEVARFGLDYSEDIAVOQMSVDQLPESFRLET  
GLESVTRTLPYNGSVVPOWLTSDGVRVCTGYSALGITSNPEFLRTATLAR  
FDEIETVTSGLPASPVDNRVLDVFPNNIILPGCAAVTHHGAGSMATALHGVQ  
ISVAHMDCLRGORTAEIAGGVFLAPDEVDATLWQALATVVEDSHAEAKELRE  
ALAPTPAEVPLTEALAHORADR"  
complement(10483..11424)  
/note="unnamed protein product; megBIV, TDP-hexose  
4-ketoreductase, TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 12= translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37815.1"  
/db\_xref="GI:13938936"  
/translation="MTRHVTLLGVSGFVSALIREFTTFLRLRAVARTGRDPPGS  
AGIEHLRVDLBEPRAVOVADVDVVLVAAGSITMSAATVPAERVNAQIMD  
LVNLAARBPAPVILFASITQANPAAPRVAOHTKAEARILROATEDGVNDVILR  
LPAITGHSGSGCTGCGVVTAMLRPALAGEPITTMHSGSRRLNLTHTVATFTAL  
LNHEALVGDVWTSSAUBARPLGIFETVAASVAKQTGNPAVPVAVSPPEANAANDR  
SDPDFSTEFRTLGMHPRVPLAGIDRTVALI,STKE"  
12181..22821  
/note="unnamed protein product; megAI; SEQ ID NO: 13=  
translated amino acid sequence"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CAC37816.1"  
/db\_xref="GI:13938937"  
/translation="MVDVPLDGRTPHPPLPMPPLCGHNEBELARAROLHAYLE  
GISDDVVAAGALAEETRAODPRAVAVASVTELTALALAOGRPHSVYVGA  
RPTAPVTVLPQGAQMPKATRLIASEPVAANMRACTRAAPREVTMSLTPYVGRVA  
HLRREVEVQALRAVQVSLALNRSFVRDADALGHSIGLIALAEVGAQDVEAARA  
AALMSREMLVGRGMAAVALPAELAEVVERDDVAVGAPSPSVLLTGAPEE1  
ARRVAELIAQVAAQVNVNMAAHSAAQVDAVASEGRMDTFVAGSDVYVYAGLTG  
RLDTRBELGADHMPRSFLPVRFDEATVATVAGTGTFTESPHVLAAISLQOQTDVYG  
SPAIVPTLDRDQGLRRLIYAAVATVGVTDVMTAAVPGVTPGHLSAIVAATDDEG  
PSTFEDMAFDVHVARLRLITVGAETALAGREVDATATRELDVLAVOLATRLA  
TAGRGDHLIMLDHPHALTEALNGLGPEERSGRSTHPTAEDEPRAVAMAC  
RLPGGVSPSEFMELEAGDAVGLPTDGMWDLDSFHPDPTSGPRAHQAPAGFLTG  
ATSPDAFPGLSPEALAVPQDRTLELSEVLEAGIPTSLRTSGTGFVGLIIPQ  
EYGPRLAEGSGVAGTGTTSVAGRAVATLGGPAISVDTACSSSLVAHLAC  
QSLRGSSTALAGVTVMPTPMGLVDFSRMNSIAPRGSAPSAADRGEMARGAGM  
LLERLSDARHGHVLAIVRGTVANSDASNGISAPNGRAQVAVITQALLASGLTPIH

CDS

CDS

Query Match 100.0%; Score 601; DB 6; Length 47981;  
Beet Local Similarity 100.0%; Pred. No. 6,5e-52;  
Matches 601; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCTCTTACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCCCTGCTGCTGCGCGCAGT 60  
DB 37506 GCTTCTTACGCGCGGCGCAACGCTTCTCTGACGCTTCCGCCCTGCTGCTGCGCGCAGT 37565  
QY 61 GGGCTGCGCGTCACTGATGCTGCTGCGGCTGTGGGCGGGGGAACAATGAGCGGCTAC 120  
DB 37566 GGGCTGCGCGTCACTGATGCTGCTGCGGCTGTGGGCGGGGGAACAATGAGCGGCTAC 37625  
QY 121 GAGGCGCGGCACTACTGCGGACGCAAGGCGCTGCGCGCATGAGCCGCGAAGCGGCGCATC 180  
DB 37626 GAGGCGCGGCACTACTGCGGACGCAAGGCGCTGCGCGCATGAGCCGCGAAGCGGCGCATC 37685  
QY 181 GAGGAGCTGCGGACCACTCTGAGCGCGGGGACCCGTGGGTGTGCTGCTGTGACCTGAC 240  
DB 37686 GAGGAGCTGCGGACCACTCTGAGCGCGGGGACCCGTGGGTGTGCTGCTGTGACCTGAC 37745  
QY 241 CGGAGCGGCTTCGTGCAACTGTTGACCGCGCGCGCGCGCGCGCCCTTGTGAGCAATC 300  
DB 37746 CGGAGCGGCTTCGTGCAACTGTTGACCGCGCGCGCGCGCGCGCCCTTGTGAGCAATC 37805

Qy	301	GGTGGGGTCCGCGCGCGGGGCGAGAGAACCGGTCAGGAAATCGAATCTTCGCGCGGGCTG	360
Db	37806	GGTGGGGTCCGCGCGCGGGGCGAGAGAACCGGTCAGGAAATCGAATCTTCGCGCGGGCTG	37865
Qy	361	GCGTCGATGCGCGAGGCGCGAAGTCGACGAGCATGTCGCGCGCTGTGTCGAGCGCGAGTG	420
Db	37866	GCGTCGATGCGCGAGGCGCGAAGTCGACGAGCATGTCGCGCGCTGTGTCGAGCGCGAGTG	37925
Qy	421	GCAAGCGGTGCTGGGCGCACCGGCGACGCGCGATGACGCTGAAGCTGAAGCTTCGCGTGAAC	480
Db	37926	GCAAGCGGTGCTGGGCGCACCGGCGACGCGCGATGACGCTGAAGCTGAAGCTTCGCGTGAAC	37985
Qy	481	CTGGGATTCGACCTCATGACCGCGCGTGAAGCTGCGGAAACCGGCTCGCGCGCGTGAACCGGG	540
Db	37986	CTGGGATTCGACCTCATGACCGCGCGTGAAGCTGCGGAAACCGGCTCGCGCGCGTGAACCGGG	38045
Qy	541	GTCCGGGTGGCGACGACCATGCTCTTCGACCAACCGGACAGTGAACCGGCTCACCGCGCAC	600
Db	38046	GTCCGGGTGGCGACGACCATGCTCTTCGACCAACCGGACAGTGAACCGGCTCACCGCGCAC	38105
Qy	601	T	601
Db	38106	T	38106
RESULT 3			
SERERYAB		20235 bp	DNA
LOCUS			linear
DEFINITION			BCT 26-APR-1993
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
COMMENT			
FEATURES			
source			
gene			
CDS			

LSGSGMAAALGSAVREBLRPMDLSTGLVAANPGRSVVSGEPALAAFSBDCAB  
GTRPADIDVYASHSPQIERABEYBLLETTGDIAPRPARVTFHSVYSKMDLELDLR  
YMYRNLRTVPRPADVNTLBAEYBLLETTGDIAPRPARVTFHSVYSKMDLELDLR  
LHRDGDLSAPLRSMATVHVSVDTRMDVLAALGAPALPTPTPQKRWLQDAAPAA  
ASDELDAVHVSMTPIREPSGNDGMDLVTPLEISPMETMCEAIYANGKWLARCEDV  
TSASVTEMAVVAQKGTGFRGLTSLSPMSCEGVPAGVGLITLVAADGADPA  
PVYCTTGCATVPRPADDLARPQTTNAGRAQVAGLEPERKMGVYDILPSVDAALRL  
LVATLGGGRADHDLVDRGRLHGRVVRASLPQSGSRVTHPGVTLVTGASVFGD  
LVPRMLADGAERLVLAAGCPGDDLLAAVEAASAVNCAQDAALBALEBDEVTVAL  
HAGTUTGGLTSEVVEPBEARLEIAKTLALVDEYVAGRAVEREYCSGASVWGGA  
GMAAAYASAYLDALAEHHRARGRESTSVAMTPMALPGCAVDGGRGRLSTLSDR  
MRMTREYTLAAGPVAAVADVMWPTLSEGFATRPALPAELLAGRGCAEAEPDSGT  
GERPORAAGLSPDQENMLBELVANAVALVEGHSAAETVYRASPSELGDSLNNML  
KRKASVGLRLPASPVLVDHPVTLAQLRLRLVGDAAQAVRVVGADESPELVIG  
GIGCFPGGIGSPPEBVLVDEHNGVANTGFPRLKMDIGQLVHPDNDGTSYVDKAG  
VGIKTVLMRMDGLPRSLHABELSHIIMESGASVEYLRVEVPACBRRPRADSVFG  
VSTVATVYIEBAPLOEARATERPRLTFLPSGRSAEVLVAADAPALAEHLRDEPL  
TDAAATVTLAAGRPAVRVAAVYEDBSGACVCAELDALKEGPSADVAVPYSAPEKVLV  
RPGCAQVYGMARDLSESSVYBASMCSKCALSHPTMKLIDVYRGDGPREDHVR  
VPGQVLESIMVSLAEHAGVTAAMVNGSGEPLAAHVAALNPGTSALEAKVHALNSQ  
VLRLELDADGASVVASDELETVLARVMDGRVAAVANNPGTSVVAAPXALDEFPAL  
AABEMKRRRLIAYRVAASPEVAREIEDRLAELGTTIYARGSVPLHSGVYAEVIDSA  
MDASVYVYVLEPFOAVYGLVEGQPTFVEYSPHPLVLAABETAHGAEMVVC  
PTLRGSGHPHLEPLNLRVHVGAGDLPRAVAGGRPAELPTYPREHOPMPRPHR  
ADVSAVLTGRGAHRLYLLAANDVPHGCAVFTGRSLSTDEQPYLAENVGRTVLPGSVL  
VDLAAAGSDGLVPLBELVQRPRLVYLAGAGALNMSVAPBESGRTIDVHAABEQV  
VLAAGSQHATGTLGAASVAGPRTGEMPEBDAVARIPLDDHLYLARGGXEYGPFO  
ALIRAMRKDSVYAEVSLAABEGFAHPLVDLAAQTLSTLCAIGSPGGKLEPANT  
VTLHAGSATSRYVAVATPAGDAAMALRTVPDAGHLVATVDSLVRSXGEKMOEPGGG  
RGEVTLAMGMLAEBSGSRGVAAADPDLATVLSGSEPRDAVTVYRPEGDDPRAA  
RHYVLMALVYRRLDEGERLEGATVYVLTSGCAVTVSDDSVPEPRAAMVYRQCA  
ESPRFVTLDTAEFGMLPAPFNDQLALRGDVAVPRSLSPASRLPIAGQVLPV  
VGTAVIDSAVEPAPADVEQPLRAGERVAVNRAGVAFVRLALGMYPQKADMTKEAAG  
ADVVAAGVVAPEPDRVYGLFOGAFAPLAVTDHRLARVPDGMSPADAAPAAVEIATY  
LHAYLHADLAGRAGOSVLIHAAAGSGVMAALVABAGAEVATATGAPAGVMTALGD  
ADENHLSARETRGAPKRPBRRGGRGVVDVANSIGRLDESDAATLADGCVFTGRLGTD  
LKDADGPRGRVAPRDLGERGDRDLRLIARVYGLLAGAGDLDLPYSAMELSGSPALQ  
HMSKRSQKGLVLTROPADVDPGYLTIGGTSTLRLALPHRLVTEHGVNAAGVLDLGR  
HAPSRDELRABEIDBLGASABTAACTYDARDLSFALLDGLPRLHGVHNAAGVLDLGR  
VSTLIDEPABEVQELPAKALMGMLAQSSEMTSGLDRIARTGVALLPRTALALPISA  
LIRGEVGAPEPISIRNSLRSGYASGADQVRCWYRATLPAAGCAEAAGRPVRLTLAGSE  
SDQVAGVLAPELRSHAAVSGYSGADQVLEBRKARVPLDGPDSLALAVELRPLGTRTGR  
PSTVTFPHDPLVLAEAHARDRLPAASPAVDIGDRIDELKALBAEABGHDVQRL  
BSLRLRNNSRADAPSTSAISBDSDBELFSMLDORFGGBDL"

19. .4470  
/gene="erya"  
/note="putative"  
function="approximate span of module 3"

97. .1482  
/gene="erya"  
/note="putative"  
function="approximate span of beta-ketocycl ACP synthase  
of module 3"

1693. .2670  
/gene="erya"  
/note="putative"  
function="approximate span of acyltransferase domain of  
module 3"

3406. .3921  
/gene="erya"  
/note="putative"  
function="approximate span of beta-ketoreductase domain  
of module 3, possibly non-functional"

4171. .4428  
/gene="erya"  
/note="putative"  
function="approximate span of acyl carrier domain of





**Source**

```
source      1. .20235
            /organism="unknown"
            /mol_type="unassigned DNA"
origins
```

Query Match	66.2%;	Score 397.8;	DB 6;	Length 20235;
Best Local Similarity	78.9%;	Pred. No. 1.5e-31;		
Matches 474; Conservative	0;	Mismatches 127;	Indels 0;	Gaps 0;

QY	1	GCCTCTACACGCGGGGCAACGCTTCTCTGACGACCTTGCCTCGTGTGCGAGCCAGT	60
Db	14500	GCCCTCTACGCGGGGCAACGCTTCTCTGACGACCTTGCCTCGTGTGCGAGCCAGT	14555
QY	61	GGGCTGCGCGTCACTTCGATCGCTCGGAGTCTGTGGGCCGGGAGAACATGCGCCGTACC	120
Db	14560	GGGCGCGCGGTGAGTCTCATCGCTCGGAGGCTCTGTGGGCCGGGAGAACATGCGCCGGAGAC	14615
QY	121	GAGGGCGGCGCACTACCTGCGCAACCGAGGCGTCCGCGCATGGAACCCGACGCGGGGATC	180
Db	14620	GAGGGCGGCGCACTACCTGCGCAACCGAGGCGTCCGCGCATGGAACCCGACGCGGGGATC	14679
QY	181	GAGGAGCTGCGGACCAACCTGAGACGCGCGGAGAACCCGTGGGTGTCCGTTGTTGACCTGAC	240
Db	14680	GAGGAACTGCACATCACCTTCGACACAGTCAACGTCCGTCTCGTTCTGTGACATGAT	14735
QY	241	CGGGAGCGGTTTCGTCCGACTGTTTCAACGCGCGCCCGCGCGCCCTCTTTCGAGAACTC	300
Db	14740	CGCAGCGGTTTCGTCCGAGCTGTTTCAACGCGCGCCCGCGCGCTGTTTCGACGAGATC	14795
QY	301	GGTGGGGTCCGCGCGCGGAGCGGAGACCGAGTCAAGAAATCCGATCTCCGCGCGGCGCTG	360
Db	14800	GCCGGTCCCGCGCGGAGACCCCGCGAGCGAGAGAGGCGCCCGCGCTCGCCAGCCGCTC	14855
QY	361	GCGTCAATGCGCGAGCGCGAACCTCAAGACATGTCCGCCGCTGTGTCGAGCGCGAGTGT	420
Db	14860	GCGGCGCTGTCAAGCGCGGAGAGGCGCGAGCACTCGCCCACTGATTCGCGCGCGAGGTCT	14915
QY	421	GCAGCGGTGTGGGCGCCACCGGCAACGCGGACGCTGATGAGGCTGTACGTCCCTTCGCTGAC	480
Db	14920	GCCGCGGTGTCCGCGCACCGCGGACGAACGCGCGCATGACCGCGCAACCGCGCTTCCCGAC	14975
QY	481	CTGGGATTTGCACTCAATGACCGCGCTGCACTTGCGGAAACCGGCTCGCGCGGAGTACCGG	540
Db	14980	CTGGGCTTTCGACTTCATGACCGCGCTGCACTTGCGGAAACCGGCTCGCGCGGAGTACCGG	15035
QY	541	GTCGCGGTGCGCAACGACATGTTCTTTCGACCAACCGGACAGTGTGACCGCTCAACCGCGCAC	600
Db	15040	GTCGCGGAAACCGCGAGCGGTGTTCTTTCGACCAACCGGACCATCAACCGGCTCGCGGAC	15095
QY	601	T 601	
Db	15100	T 15100	

RESULT 5	AR095529	20235 bp	DNA	linear	PAT 08-SEP-2000
LOCUS	AR095529	Sequence 3 from patent US 6004787.			
DEFINITION	AR095529				
ACCESSION	AR095529.1	GI:10023473			
VERSION					
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 20235)				
AUTHORS	Katz,J., Donadio,S. and McAlpine,J.B.				
TITLE	Method of directing biosynthesis of specific polyketides				
JOURNAL	Patent: US 6004787-A 3 21-DEC-1999;				
FEATURES	Location/Qualifiers				
source	1..20235				
ORIGIN	/organism="unknown" /mol_type="unassigned DNA"				

Query Match	66.2%;	Score 397.8;	DB 6;	Length 20235;
Best Local Similarity	78.9%;	Pred. No. 1.5e-31;		
Matches 474; Conservative	0;	Mismatches 127;	Indels 0;	Gaps 0;

QY	1	GGCTCCCTA	CGCGGCGGAGCAACGCTTTCCTCGACAGCCTTGGCCCGTGTGAGTGTGGGGCAGT	60
Db	14500	GCCCTCTTA	CGCGGCGGAGCAACGCTTTCCTCGACAGCCTTGGCCCGTGTGAGTGTGGGGCAGT	14559
QY	61	GGGCTGCGCGGTCA	CTCGATTCGCTGTGGGGTCTGTGGGCCGAGACATGGCCGGTACC	120
Db	14560	GGCGGCGCGGTGACGT	CCATTCGCTGTGGGGGCTCTGGGCGGAGACATGGCCGGGAGAC	14619
QY	121	GAGGGGGGAGCATTA	CTTGTGGCAGCCAGGGGCTTGGCGCCCATATGATCCCGCAGCGGGCGATC	180
Db	14620	GAGGGGGGAGAGTA	CTTGTGGCAGCCAGGGGCTTGGCGGCCCATATGATCCCGATCTGGGCGCGTC	14679
QY	181	GAGAGCTGGGACCA	CCCTTGGACGCGGAGGACCCGTGGGTGTCTGGGTGTGACCTTGGAC	240
Db	14680	GAGAGACTGACATCA	CCCTTGGACCAAGGTCAAGAGTCCGTGTCTGGGTGTGACATGAT	14739
QY	241	CGGAGCGGTTT	CGTGCACATGTTTCAACCGCGCCCGCGCGGGCCCTCTTTCAGCAATC	300
Db	14740	CGGAGCGGTTT	CGTGCACATGTTTCAACCGCGCCCGCGCGGGCCCTCTTTCAGCAATC	14799
QY	301	GGTGGGGTCCGCGG	CGGGGCGAGGAGACCGGTTCAGGAATCGGATCTGGCCCGCGGGCGTG	360
Db	14800	GCCGGTGTCCCGGG	CGGAGACCCGGCAGACGAGAGGGGCCCGGCGCTGGCCUAGGGCTTC	14859
QY	361	GCGTCATATC	CCGAGGCGGACATGATCAGACATGTGCGCCCGGCTGGTCCGAGCCGAGGTG	420
Db	14860	GCGGCGCTGTGCA	CGGCGCGAGGCGCGAGCACTCGCCACTGATTCGCGCGGAGGTC	14919
QY	421	GCAACGCTGTGGG	CCACGCGCACCGCGACAGTGAATCGACGTGAACGTGGCTTCCGTGAC	480
Db	14920	GCGCGGCTGTGGG	CCACGCGCACCGGAGATGACCGGACCGGCGCTTCCGTGAC	14979
QY	481	CTGGGATTCGAC	CTTCATGACCGCGCTGCACTGTGGGAAACGGGCTCGCGGCGGTGACCGGG	540
Db	14980	CTCGGCTTCGAC	CTTCATGACCGCGCTGCACTGTGGGAAACGGGCTCGCGGCGGTGACCGGG	15039
QY	541	GTCGCGGTTGAC	CGACATCGTCTTTCAGACACCCGACATGTGAACCGGCTTACCGCGCAC	600
Db	15040	GTCGCGGAAAGCG	CGACCGGTGTCTTTCAGACACCCGACATGTGAACCGGCTTACCGCGCAC	15099
QY	601	T	601	
Db	15100	T	15100	

RESULT 6	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
AY661566	AY661566	Saccharopolyspora erythraea polyketide synthase gene cluster, complete sequence.	AY661566	AY661566	GI:55419414		Saccharopolyspora erythraea Saccharopolyspora erythraea Bacteria; Actinobacteria; Actinomycetales; Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.	1 (bases 1 to 32299) Kodumal,S.U., Patel,K.G., Reid,R., Menzella,H.G., Welch,M. and Santl,D.V.	Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polyketide synthase gene cluster	Proc. Natl. Acad. Sci. U.S.A. 101 (44), 15573-15578 (2004)	2 (bases 1 to 32299) Hu,Z. and Reid,R.	Direct Submission	Submitted (21-JUN-2004)	Kosan Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES  
source

gene  
CDS

```
Location/Qualifiers
1..32299
/organism="Saccharopolyspora erythraea"
/mol_type="genomic DNA"
/db_xref="taxon:1836"
1..10638
/gene="eryA1"
1..10638
/gene="eryA1"
/note="DEBS1"
/codon_start=1
/translation="1"
/product="EryA1"
/db_xref="GI:55419415"
/db_xref="GI:5541820.1"
/translation="MADLSLSDSRTNQPRIVRPWPLSGCNESALRAROLRAHLD
RPDAGVEVGAALAHDEQADGPHRAVVAASSTSELDDGLAVALPRAVYRGTA
RPSAPVVFVPGQQAQWAGAGELGESSRYAAMDCARAFEPVTDWTLAQVLDSE
QSRREVVVQALPAVQTSIALMRSFGVTPDAVVGHSIGELAAHVCAAGADAAAR
AALMSRMTPLVNGDMAVALSADIEPRIARMDDVILAGVGRSVLLTGSPEVY
ARVQELIARGVRAOVINVMASAGVDIAGCMRSALAMPAGSGSEVPYASLNGS
AVDTRELVADYMRSSFLPVRPDAITSLAEVCGTVEASPHVYASALAQTLDAAG
SSAUVVPLDRQGGMRFLAAQAFTGVAVDTVAIDVAGPEGLSEFPABEE
DEPAGSGVWNAAPHVIRELAVNGETALAGREDAETFEELGLDVLAAQLAA
KVSALIGREVNIALYDPTPRALAEALAGTEVAQRETRATTEAAGSPVAVANA
CRLPGVSTPEEFWEILSEGRDAVAGLPTDRMDLSEPHDPTRSGTAHQGGGFLT
EATPDPAPFGMSPREALAVDPOORLMLLEISWEVLEAGIPTSLSAPGVVGLIP
OEYGPRLABGGBGVGLMTGTTTSSVAGRIATYLGIRALISDTYLCSSILAVHIA
COSLRGESSILAMAGSVTVPQMLVDFSWNSLADGRCKASAGANGMGAGAG
MLLERLSDIARRNGHPLYALVLRGTAVNSDASNGLSAPNGAOVRVIOQALESGLP
ADIDVAHGIGTGLDPIEARLFEAYGDRDEQPLHGSYKNIQHTQAAAGVAYI
KMVLAVAGTGLPTLTHASERKEIDMSGAI SLIDDEPMPAGAPRPAAGSFGIG
TMAHILIEBAPOVVEGRVAGGVAPVYASASRGLRAQALAHLEHPPQDRR
DIAYSLTGRALPHRAAPAVDSALRLVLDGLATGNAGDANGTRAQORAVFVP
GQWQWQMGMAVDLIDTSPVFAALRECADLLEPHLDEVLIPTEABAREQDALS
TBRVDVQPVMAVAVWSIASMWRAGVPAVYIGHSQGIAAQVAGALSDDARVVA
LRSVIAITMPKNGKMAIIAPAGEVRARIGDRVYIAANGPRSVVAGDSDELDRVA
SCCTECIRAKRLADVASHSHVETIDGLAEIGDEPHPLGFPVFPSTVYGTQTP
DELDAGWYRNLRTVAPDAVRLAAGCYRTPLEVSANPLTALITREIGDGGADLS
AIIHLRGDGLADFGELSRAPAGVAVDMESVHLTGARVPLPYTPROREVNIE
PKPARSTEVDEVSALKRYLIMRPTGASEPARLDGWLVAKYAGTADSTABARE
ESGARVRELVDARCGRDELAERLSVGEVAGVLSLAVDEAREERAPLALADLT
LSLVQAMVASLGCPLMTVTSAAVATGFERVRAAHGALMGVRLALEMPAGGL
LVDVAGSVAEILARHLAAVVGAGAGEDOLATDGVGRWVRRAAPATDMCKGTGY
LYNGCTGVGGOIARPLARBGAPHLILVSGSPDADQALVATLEALGATVTAAD
VTDRBSYRELGGIGDVPVLSAVTHAAATLDDGVYDTITGERIRASARKVLRNIE
ELTRELDLTAFVLSSFASAFAGAGLGGYAFGNAVYLDGLAQORSGLPATVAMGNI
AGSMAEGSPVADRFRRHGVIMPEPETACRALONALDRAEVCIIVYRMRFLAYTA
ORPTPLFBEIDARRAAPOAAAPRVCAGLSIPAPERKALFELVRSILAAVCHASA
BRVPADQAFELGVDSLSLELRRLGAATGRLPTTVFPHPVRTTAAHLAEGLG
ATGAEOAPATTTAPVDEPIATVGMACTLGRVDSPERLMELITSGRDSAEVPPDRG
VPDELMASDAAGTRARHGNPAAGDPPAAFGISPREALAMDQOQALETTYALE
SAGIIPETLRGSDTGVFMKSHQCATGRPEBVGDSYLTNGTASVAGRIAYVLS
LEGALTVDTDCSSILVALHTACSLRNGDCGLVAGGVSMAGPEVETESRQALS
PDGRCKPESDADFGIGEGSAFVLQRLSDARREGRLVGVAGSANOAGANGLS
APSGVAOORVIRBAMARAGITGADVAVVEAGTGRRLDPEVASALATYKSGSGS
PVLGSKYSNIGHAQAAGVAGVYKVLGLERGVPPMLCYGERSGLIDMSGGIELA
DGVKMSPADGVRACGATGAGVSGTNAHYITIAEPPEPFPQRRMLPATGVVPL
SARTGAALRAQAGRLADHLAHPGIAPDVSWTAKARQHEERAAVLAATGVAVR
LRAVADGAVVPGVVTGASDGSVFVPGQAOWMGARBLIPVPFASVLEICDAVL
SEVAGFSVEVLEPPADPESLERVVQVPLFAVWSILARLMACGAPASVAGISQV
ELAAVVVAGALSLEDMGVVARBSRAVARAVAGSRMSLVGRGSRVVEGLLDDSGTR
LEVAANVPDAVVVAGDQAAREPLFYCBGICIRARAI PUDVASTHAEVVRBEVIO
ALAGITPRRAVPEFFSTLQGLDSTEDLAGWTRNLHPVEFSAVALTDQGYATF
IESPHVPLVASVQETLDDASDAVLTGLTRDADADRFLTALDATTRAVADWMEA
VLSPAGLVDLPCEYPOGGRFMLIPDRTPRDELDDMFYRVDTTEVPRESEPLAGRWL
VVVPGHEDEGWTVEVRGALAEAGAPETVGVGLVGDCAVVSILALEBDQAQVLT
VIVRELDARGIDAPLWTVYRGAVDAGSPVARPDQAKMGIGQVASILRKGPMWGLYD
PHMDPBLRGRITAVLASBOVAVRADVARRLSPAHYTRTSYAVPGCTIYVTCG
TAGLGAERWMLAGRGAEHLALVSRGPTDGVGLTAELTRLGARVVAHCDVSSRE
FVRLEVHGLIEBGDVAVGVNAAGLPOOVAINDMDEAFLDVAAKAGAAVLDLBSL
DAELFLPSSGAGVGSARQGYAAGNAFLDAFAHRRGRGIPATSVAMGMAGMGMT
GDEAVSFLRERGVAMVPVPAALADLVILASGETAVVVVTDVMDPAFASVTAAPRP
```

gene  
CDS

```
LIDRLVTTASERBAGEPTESTLDRLAGLPRARTEALVRLVTSYATVYLGHDDPKAV
RATTPKEIGFDSILAAVRLRNILNLAATGLRLPTTLFVDHNMASAVGFLDAEIGTER
GEAPSLAGLDLAEALPREVPATERRELVQRLERMLAALRPVQAADASGTGANPSGD
DYGAGVDLELRLRGLELDGD"
12080..22783
/gene="eryA11"
12080..22783
/gene="eryA11"
/note="DEBS2"
/codon_start=1
/translation="1"
/product="EryA11"
/db_xref="GI:55419416"
/db_xref="GI:5541821.1"
/translation="MTDSEKVAEYLRRAATLDLRARORIRELSDPIAYVMAICRLPG
GVNTPQRIEMLIARGSETLSGFTDGMGLARLHPDPDRGTSYVDKGFGLDAAAF
DAFPQVSPREAAAMPQORLLLETSWEIYENAGIDPHSLRGATGVGFLGAAPGGE
DTAAEDVBSITGTVAPASARISTYWGLEBPSSISVDTRACSSILYALHIAESLRK
GESSMVVGGAAMATPGVVDSEKQRLAADRSKAPAGAGAGFGFBSGVTVLLEBR
LSRARRNGEVLAVVSGSLAMDQASNGLSAPGEPQORVIRQALBSQGLEPDVDVAV
BAHGTGALDGPTEANALDQYGRDDADPLMFGSYKSNIGHTQAAAGVGLIKVTL
ALANGRLPATLVREBPTPHVDMSSGGVALLAQNQPMRGRGRTAAVSAAGISGTNAH
VYBEAPERHRTTHADGRPVLYVUSABETALLRQAQIATLEBRPDADLGVGIG
LATTARAHERRAAVASTREAVRGLREILAGAAITADAVTEGTTEDGRVYVLPFGQ
LSQWAGMGAEILSSVPFAGKIRACESMAMPQDWKVSJDLRQAPGAGJDRVDVQV
VLEAVWSLAEILMRSYGVSEBPAVVVGSQGIILAAHVAGLTLDDAALVYGRSLMRS
LISEGGMAAVALDEGAAYREBLRPMODELSVAAVNGSVVSGSEPCALRAFSBECAR
GIRVROIDVNYASHSPQIIBERVREILLETTDILPRAPVTFHSVTSRSMQDLEDR
TYRNLRETYRPMADATYRLABSGYDAFIVSPHPVYVQVBEAVERADQADEVVVGS
LHRDQGLSFLMSMAHAVSGVDIRNDVAPGAAPALPTYTPQORRYTLQAPAPAA
ASBELVAYSVMTIIEKESGDLQDMVLVPLISPEVTEMLCEATINANGRLACSDY
TSASRTEMAQVAVOAGTGFSGVLSLSDSACRPGVPAGAVGLTLVQALGDAGVDA
PMWLTQCAVTPADDDIARPAQTTAGFPAQVAGELPGMWGVLDPEESVDALALIL
LVAVIRGSGRAEDHILVRDGLRHGRVRYRSLPOSGRSMTPGVTVGTAAASPVGQ
LYRMADRGAERLVLAGACRQDDLLAVEREAGSAVYCAQDAALAREALDREPTALY
HAGLTVNGSILSEAPFERMETIATAKTALLALVDEVTGDAVEREYCSSVAGIWGCA
GMAAYLNGSISEAPFERMETIATAKTALLALVDEVTGDAVEREYCSSVAGIWGCA
AMPTWERYVLAAGSVSVAVADVPVSGEPAATPTALFAELARGQQAERPDGPT
GAPQORLAGSPBEOQENLELVANAAYVYIGHSASLEINVRRAVSEFLGDSILAMAL
RKULASTGRLPASIVFHDPTVYALQALRALVGDADQAAVVGGAADBSSEPIAY
GICRFPFGGIGPEQLRVYALBQANLTTGPPLDRBMQMGITLHYDDPNPGTSYVDKXG
FLTDADPFGPFOITRYBRALANDPOORLMLEKMEAVEBAQIDPDRLRTDTGVFPG
MNOQSVYQLAGEAERVDYQGLSNSVLSIGRAYFMEGEPALVYATDCSSLVGL
HLMQALRREGESLALGATVWSPTPTDPTFQORLMDLVDVRDGGDPHERVD
GVVALVLEPSRANNGHOVLAIRGAVANGDSNLELAPNGSQORVIRQALAAAG
VPADVYVEPSRANGTELETPITRAGALVAYQDORDEPLGSKYKNIQHTQAAAGAA
GVYKTVLAMEHGMPLPRSLADELSPHIDMSGANBYREVPVAGRGPBRAGVSSNG
VSGTNAVITVIEARAEQEAARTSGPLPEVLSGSESVVANAQRBALBHLRDTPELIL
TDAAMTATARARPDVRAAVLGDRAVCALDGLAAGRSADVAAPVTSAPRPLV
FPQQAQWQMGARBLLESSSVFASMSGRCAELSPTMDKLLDVRDGGDPHERVD
VLOPVLFSIWSLAEILMRAGVTPAAVVGHSQGIILAAHVAGLTLBAAAVVLSQ
VIRELDQGMVSVGASRDELTVALAMDVRVAVNAANGPSTVAGPTAELEDFEFA
ABAREMPPRIATVYASHSPREYIIRLAAELGTTITANGSVVLHSTVYGRVLDISA
MDASYRNLRLRPVLEQARGLVQEGFTDFVEYSPIPLIMAVEYTEBAHGAATVCY
PTLARSGPHEFELRNILRAHVHGVADLRLPAGSGAPLEPLTYPEHQRWPEPHRP
ADVALSGRGAEHLILAAVDPDEGPEMLHAGPELHVVVGRRLVGLVGSVL
VDLALAGEVDGLVELEVLQORPLVAGAGALRMSGVAPDESGRRTIIVHAEADVA
DLDADQMSQATGTLAAGVAGAPRDTQWPEBDAVRIPLDHYOGLABOQYEGVSPFO
ALRAWRKDSVYAVESILADEBSYATHPLVDVAPQTLISGLLGEQGGGLPRAMNT
VITHASGATSVRAVATPAGDAAMRLTDPHGLVAVVDSLVASTSGEKKQPEPRGG
EGEHLIASRETFGRKRBRGRGVGVNVNLSLGEILDESDALADGAVENAGKTD
RHGVTLMALVRLRLQEDELGATLVYATSGAVVSDSDSPBECAAMAMGVITCAQV
ESPPRFLVLDTDAREGMLPVPNDROLARDQDVFVRLSPTLASALTLPGQTVLRP
GDGAIDVAPERAPADVSNQPLRAGEVRDYVATGNPNFDVLLALGYOKALMGTEAAG
VYIVAVGPDVADGADVRLFOQAPAPIATVDRKLARPVDGSDDAAPVPIAYTT
AHVLAHLDAIGLRAQOSVLIHAAGGVGAAPALARRKAGVLAIRGPKGTLLAAGL
DDEHIIASRETFGRKRBRGRGVGVNVNLSLGEILDESDALADGAVENAGKTD
LHDGDFRGRYAPDLGAGADRLGSELLEVGLLAGEILDELDRLSASWELGSAADALQ
HMSRGHVVGLITOPAPVDPDGVTLITGGGTGLGRLLARLVTHEGRLHLLVSRG
ADAGSBELEKATIEDGASATIAACOTPDADRALALALDDGLPRPLTVGNVHAAGVLDGL
VTSIDEPAVEVEVLAKVDAAANLLELANTGLSFFVLPSASAISTLAGGQGVYAAAN
SLNLAALVLRRTGIPALALGMAQASSEMSTSGIDRLARTGALALPTEPRALAFDSA
LRRGEVVFPLISIRSALIRAEFVEVYLRGMVRAKTLRAAQAEAAAPVVDRLAGRSS
```



mat\_peptide

57..9590

/gene="eryA"  
/product="6-deoxyerythronolide B synthase"

ORIGIN

Query Match

63.0%; Score 378.4; DB 1; Length 9769;

Beet Local Similarity 77.8%; Pred. No. 1.6e-29;  
Matches 470; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

```
QY 1 GCGCTCTACGCGGCGGCAAGCCCTCTCTGACGCGCTTGGCCGCTGCTGTCGCGGCGAGT 60
DB 3658 GCGCTCTACGCGGCGGCGGCAAGCCCTCTCTGACGCGCTTGGCCGCGGCGGCGGCGAG 3917
QY 61 GCGGCTGCGGCTCACTCGATGCGCTGCGGCTGCTGCGGCGGCGGCAAGCATGCGCGGTACC 120
DB 3918 GCGGCGGCGGCTCACTCGATGCGCTGCGGCTGCTGCGGCGGCGGCAAGCATGCGCGGAGAC 3977
QY 121 GAGGCGGCGGCACTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 3978 GAGGCGGCGGCACTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4037
QY 181 GAGGAGCTGCGGAGCAAGCCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 4038 GAGGAGCTGCGGAGCAAGCCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4097
QY 241 GCGGAGCGGCTTCTGCGAAGCTGTCACCGCGCGCGCGCGCGCGCGCGCGCTTCTGCGAAGACTC 300
DB 4098 GCGGAGCGGCTTCTGCGAAGCTGTCACCGCGCGCGCGCGCGCGCGCGCGCTTCTGCGAAGACTC 4157
QY 301 GGTGGGGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 4158 GGTGGGGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4217
QY 361 GCGTGCATGCGCGGAGCGGAGCGTCA---CGAGCATGTCGCGCGGCTGGTCCGAGCGGAG 417
DB 4218 GCGGCGGCTGCTGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 4277
QY 418 GTGGGAGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
DB 4278 GTGGGAGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4337
```

QY	478	GACCTGGGATTGCACTCCATGACCGCGGCTGCACTGCGGAAACCGGCTCGCGCGGTGACC	537
DB	4338	GACCTGGGATTGCACTCCATGACCGCGGCTGCACTGCGGAAACCGGCTCGCGCGGTGACC	4397
QY	538	GCGGTCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	597
DB	4398	GCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	4457
QY	598	CACCT 601	
DB	4458	CACCT 4461	

RESULT 8  
SERVYABS  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT  
FEATURES  
source

1. 20444 bp DNA linear BCT 28-FEB-1992  
S. erythraea eryA gene for 6-deoxyerythronolide B synthase II & III.  
X62569  
X62569.1 GI:46977  
6-deoxyerythronolide B synthase; acyl carrier protein;  
acyltransferase; beta-ketoacyl synthase; dehydratase; enoyl  
reductase; ketoreductase; multifunctional subunits; thioesterase.  
Saccharopolyspora erythraea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.

1. Bevilacqua, D.J., Cortez, J., Haydock, S.F. and Leadley, P.F.  
6-deoxyerythronolide B from S. erythraea: cloning of the structural  
gene, sequence analysis and inferred domain structure of the  
multifunctional enzyme  
Eur. J. Biochem.  
2 (bases 1 to 20444)  
Bevilacqua, D.J.  
Direct Submission  
Submitted (19-SEP-1991) D.J. Bevilacqua, Biochemistry Dept, Univ of  
Cambridge, Tennis Court Rd, Cambridge CB2 1QW, UK  
For related sequences see X56107 & M63677.  
Location/Qualifiers  
1..20444  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRRL 2338"  
/db\_xref="taxon:1836"  
/clone="SB55, DE10, PBK25"  
/clone\_lib="genomic in lambda EMB3"  
37..42  
46..20268  
/gene="eryA"  
46..10749  
/note="ORF B"  
/codon\_start=1  
/transl\_table=1  
/product="6-deoxyerythronolide B synthase II"  
/protein\_id="CA44448.1"  
/db\_xref="GI:581651"  
/db\_xref="GO:003132"  
/db\_xref="UniProt/Swiss-Prot:Q03132"  
/translation="MTDSKVAEYLRATLRLARQRIRESLSDPILVSMCLPQ  
GNTFQRLMELREGSGTSGFPDRMELRLHPPDNGTSTYVKGFLDAGF  
DAEPFGPREAAMDPOORLLRTSMELVENAOIDPHSLGTATGYLGAKEGYS  
DTAAEDVEGYSVTGVAIDAVASGISTMGLGSGISVDTCSGLVLAIVSLRK  
GESSMAVVGAAVATGVPFDPDRQALADGSKRGAGADRGSEGVTLVLER  
LSEARRNGHEVLAIVRGSLAQDASNGLSAPSGAQRVTRQLBSGSLPQVDAY  
EAHGTGALGDPIDANALDDTYGDRADRLMLGASVSNIGHTQAAGVGLNVIL  
ALRNGELPATLVEEPTPHVWSSGVALAGNOPMRGERTRAAVASFGSTNAH  
VIVEAPREHERETTAGDGPVPLVVSASTAALRAQAQIAELERDADLAGVGLG  
LATTRAHERHAAVAVASTREAVRGLEIAGATVAVVGTREVDNRVPLPQGO  
GSQAMGAKELISSPYFAGKIRACDSMAAMQMKVSDVLRQARGAGLDRVNVOP  
VLFVAVSLAEIKMSYGVPEPAVVGSHGELAAAHVAGALTUEDNAKIVVRSRLMS  
LSGCGAAVAVLGSAAVBERLRPMQDRLSVAIVNGPRSVVSGEPGALRAFSEDAAB



GIKVRDIDVDYASHSPQIERKREBELLETGDIAPRPAVTFHSTVSRSMDGTELDAR  
YMYRNRETFRPADAVYRLAESGDIAFIEVSPHPVQVAEEAEADGMEADAVVVS  
LHRDGDGLSAPFSLMATAVRLAESGVDIEMDVLPAPAAVLPPTYFORRYMYLADAVVVS  
ASDELAIVSWTPIKRESGNLDGDMVVTPLISPEKTEMLCEAIINANGRALRCYD  
TSASRTEMAQAVAAQAGCPFGVSTLSSDSASCPRPVPAVAGILITVQALGACATDA  
PWCILTOGAVRTPADDILAPQOTTHAGFQVAGLELPGMWGVVDLPESDVDAALRL  
LVAVLRGGGAEHDLAVRDRGLHGRVVRVSLPQSGRSWTPHGTVLVTCAAPVPGQ  
LVYMLADRGAEHDLVLAGACPGDILAAVEAGASAVVCAADAAALREALDEPVTLAV  
HAGTLNFGSISEVAPBEPAETIAAKTALLAVDEVGDAVEREYCSVAGIWMGA  
GMAAAYAGSYLIDALAEHHRARGSCSVMTWMLPGAVVDGTYLREBLRDLSDR  
AMTRERIVLAAGVAVAVADVDPVYSECPATRPITALPHELAGGQABEDSGT  
GEPQAQIGLSPDEQENLLELVANAEEVLGHSAAEINVRASFELGIDSLANAL  
RRLSASTIGRLPASLVFDPHTVLANAOHRLVLPADQAAVAVGADDESPIAIY  
GICGRFPGIGSPGOLMRVLAEGANLITGPARGMDIGLHYDPDNPTGSYVDGQ  
FLTDADPDGFGFQITPRBALMDPOORLMLETAMRABERAGIDPDALRDTGVVVG  
MNGOSYMLLAGABERVDGQIGNSASVLSGRITATTCBEGPALTVDTRCSSLVTI  
HLMQALRKEGCLALAGVTVMSDPTTFVSTQKLSADGCKASADPALS  
GVALVLEPLSRARANGHOVLAVLRSAVNOGASNGLAAPNGSERVIRQALASG  
VPADVDVVAEHGTETELDPIBAGALITAYGQDRDRPLTSGKTNIGHTQAAAGA  
GVIXVILAMRHGMLPRSLHDELSPHIDMSGAVLEVEEVPMPAGRPBAGVSSRG  
VSGTNAHVILEEAPABOBAVTERGRLPYLSGRSAVVAQABALAEHLRDTPELT  
TDAWTLATGRAPVYRAAVITGDRAGVCABDLAEGRSADAVAPTSAPKPYLV  
FPGQAGWGMABDLLESSVFAESMSRCAEALPHTDWLADVVRGDXGDPDEHVD  
VLQPLFESIMVSLAEMLRAHGVTPAAVNGHSGEIAAHVAGLSLEAAKVVALLSQ  
VLRELDQGMVSVGASRDELFTVLARMQGVVAAVNGPSTVAGPTAELDEFPAB  
AEAREMKPRILAVRYASHSEPARIEDRLAEELGTITAVGAVPLHSTAVGEYIDTSA  
MEASRYMRLRRPVLPEQAVRGVVBGPDPTFVSPHPVILAMVEETAEHGAERYVC  
PTLRROSGPHERLRLILAHYGVQADLAPVAGRPALPTYPPEHOFMRPHRP  
ADVSLAGREHRLLLAADVPHGGAFTGKLSDBOEWMLAEHVAGTILVPSVL  
VDLALAGDEHGLPVLBEILOPVLVLAGALRMSVGAPEDSGRITIDVHAEBVA  
DLADAMOSHATGTLAQVAGAPRDEOMPEDAVRIPLDDHDGLAEQGEYGPFO  
ALAAWRKDSVVAEVSIADEBEGVAFHPLLDVAVQTLISALGEPGGKLEPPANT  
VTLHAGATSVRVATPAGADAMALVTDOPAGLVAITDSIVVRSSTGEKDEPPEAG  
KEEHLDMVRLAEPCSTGRVVAADSDIDAVRSRPPEDAVLYRYEPEGDPRAA  
RISCTMAALVYRMILQEBELPGATLVATISGATVSDDSVPEPAAAMVTRCAQA  
ESRDRFVLLDIDEPAPMLPAVPNPQALARGDVFVRPLSPLASLTLTPAGTORLP  
GCGALDSVABEPADVEOPLRAGEVRVATAGVNRDVLALAMTPOKADMTZAG  
VTVAVGPVDAPFAPGDRVGLFQGAFAPIAVTDHRLARVPGMSDADAANAAYIAATT  
AHYALHDLAGIAGOSVLIHAAAGVGAVALVABRAGAVALATAGPAGHTLRACLT  
DDEHTASRETPARKFRRTGTGRGVTVLNTLTGELIDBSADLLEDEGVFWMKTD  
LNDAGFRKRAFPDIEGADDLSILBEVGLLAGABLDRLPSAMELSGAPALQ  
HMSRGNHVKLVITOPAPVDPGTIVITGCTGLRLANHLTEHGVHLLLVSRG  
ADAPGDELEAEIEDLGASAEIACOTDRDALSLDGLPRLTGVVHAAGVLAAGL  
VTSIDEPAYEOVLRKAVDAAMNHLTANTGSLFVLPSSASVLAGPQGVYAAANE  
SLNALALERTGLPAKALKGCMQAQSEMTSGLRIRKATGVAALPTERALALPESA  
LIRGGSVTRPPLSINRSALRAEFVPEVLKGMVAKRLAQAQAAQAPNVDRLAGSSE  
SDQVAGIAELVSHHAASVSGSADQLPERKAKDGFDSLAVEHLNRILGATVGRLL  
PSTLVDFHPTPLVAEHLDRLPAAAPAVDIDGRIDELKALAEALSAEDGHVDVQRL  
ESILRRMNSRRADAPSTSAISEDADDELFSMDLQRFGGGEDL"  
139.1416  
misc\_feature  
/gene="erya"  
/product="beta-ketocoyl synthase"  
1471.2763  
misc\_feature  
/gene="erya"  
/product="acyltransferase"  
4261.4455  
misc\_feature  
/gene="erya"  
/product="acylcarrier protein"  
4513.5787  
misc\_feature  
/gene="erya"  
/product="beta-ketocoyl synthase"  
5839.7116  
misc\_feature  
/gene="erya"  
/product="acyltitransferase"  
7117.7563  
misc\_feature  
/gene="erya"  
/product="dehydratase"  
8743.9234  
misc\_feature  
/gene="erya"  
/product="enoylreductase"  
9418.10164  
misc\_feature  
/product="ketoreductase"

misc\_feature  
CDS

10315..10509  
/gene="erya"  
/product="acylcarrier protein"  
10750..20268  
/gene="erya"  
/note="ORF A"  
/codon\_start=1  
/transl\_table=11  
/product="6-deoxyerythronolide B synthase III"  
/protein\_id="CAA44449.1"  
/db\_xref="GI:46979"  
/db\_xref="UniProt/Swiss-Prot:Q03133"  
/db\_xref="GOA:Q03133"

misc\_feature

10867..12111  
/gene="erya"

Query Match 63.0%; Score 378.4; DB 1; Length 20444;  
Best Local Similarity 77.8%; Pred. No. 1.3e-29;



Matches 470; Conservative 0; Mismatches 131; Indels 3; Gaps 1;

Qy	1	GCCTTCACGCGCGCGGCAACGCGCTTCCTGACAGCGCTTCGCGCGCTGCGCGAGT	60
Db	14533	GCTCTTCACGCGCGGCGCAACGCGCTTCCTGACAGCGCTTCGCGCGCGCGCGCGAGC	14592
Qy	61	GCGCTTCGCGCTTCACCTTCGATCGCTCGGCGTCTGTGTGGACCGGCGAACAATGCGCGTAC	120
Db	14593	GCGCGCGCGGCGGATGATCATCGCTCGGCGGCGCTCTGTGGCGCGGCGAGAACATGCGCGGAGAC	14652
Qy	121	GAGGCGCGGCGCATCTCTTCGCGACGCGAGGCGCTTCGCGCGCATATGACCCCGGAGCGGCGATC	180
Db	14653	GAGGCGCGGCGAGTACTCTTCGCGACGCGAGGCGCTTCGCGCGCATATGACCCCGGATCGGCGCGT	14712
Qy	181	GAGAGCTGCGGACCAACCTTCGACGCGCGGAGACCCGCTGGGTGTCTGTGTGTGACCTTGAC	240
Db	14713	GAGAGACTGCAACATCAACCTTCGACGACGCGTCAACGCTCGGTGTCTGTGTGTGACATATGAT	14772
Qy	241	CGGAGCGCGGTTCGTTCGAACTGTTTCAACGCGCGCGCGCGCGCGCGCTTCCTTCGACGATC	300
Db	14773	CGCAGCGCGGTTCGTTCGAGCTGTTTCAACGCGCGCGCGCGCGCGCGCTTCCTTCGACGATC	14832
Qy	301	GCTGCGGCGTTCGCGCGCGCGCGCGCGAGACCGCGTCAAGAAATCGGATTCGCGCGCGCGCGT	360
Db	14833	GCGGTGCGCGCGCGCGCGCGCGCGCGAGACCGCGCGCGCGCGCGCGCGCTTCGCGCGCGCTTC	14892
Qy	361	GCGTTCGATGCGCGAGCGCGCAACGTCA---CGACGATGTGCGCGCGCGTGTTCGAGCGGAG	417
Db	14893	GCGCGCGTGTCTGTGCGAGCGGCGCGAGCGCGGACCATCGCGCCACTGATTCGCGCGGAG	14952
Qy	418	GTCGCAACGCGTGTCTGTGGGCGCACCGGACCGCGACCGGTGATGAGCGTGAACGTCTTCGCT	477
Db	14953	GTCGCGCGCGGTGTCTGTGGGCGCACCGGAGCGAGCGCGGATGAGCGCGACCGCGCGCTTCGCG	15012
Qy	478	GACCTGGGATTCGACTCGATCGACCGCGCGTGTGACCTTCGAGAACCGGCTCGCGCGGTGACC	537
Db	15013	GACCTCGGCTTCGACTCGATCGACCGCGCGTGTGACCTTCGAGAACCGGCTTCGCGCGGTGACC	15072
Qy	538	GCGGTCGCGGCTGCGCGACGACCATGCTTCTTTCGACCAACCGGACGATGACCGGCTCAACCGCG	597
Db	15073	GCGGTCGCGGACCGGACCGGACCGGTGTCTTTCGACCAACCGGACCATGACCGGCTCGCGGAC	15132
Qy	598	CACT 601	
Db	15133	CACT 15136	
RESULT 9			
LOCUS	AY771999	31531 bp	DNA
DEFINITION	Synthetic construct DBBS polypeptide synthase gene cluster, complete sequence.		
ACCESSION	AY771999		
VERSION	AY771999.1	GI:54778584	
KEYWORDS	synthetic construct		
SOURCE	synthetic construct		
ORGANISM	other sequences; artificial sequences.		
REFERENCE	1 (bases 1 to 31531)		
AUTHORS	Kodumal,S.L., Patel,K.G., Reid,R., Menzella,H.G., Welch,M. and Santi,D.V.		
TITLE	Total synthesis of long DNA sequences: Synthesis of a contiguous 32-kb polypeptide synthase gene cluster		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 101 (44), 15573-15578 (2004)		
REFERENCE	2 (bases 1 to 31531)		
AUTHORS	Reid,R.C., Kodumal,S.J., Patel,K.G., Menzella,H.G., Welch,M. and Santi,D.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-OCT-2004) Korean Biosciences, 3832 Bay Center Place, Hayward, CA 94545, USA		
FEATURES	Location/Qualifiers		
source	1..31531		

ALAAGTERRAEVPEFSTLTGDFLDGTDLNAGYWRNLRHPEVSHAVOALTDGCVATF  
LEVPSPVLASVOETLDDAESDAAVLGTLERDAGDADRLTALADHTEGVNDMA  
VLGRAGVLDLPGYPOGKRFMLQDRTTPEDELDMGYRVDWTEVPSEBAALRGML  
VVVPEGHEDGWTWVRSALAEAGABRYTRGCVGLVDCAGVSVLLALEGDANOTL  
VLVRELDABEGIDAPLMTVTTGAVDAGSPVAPPOAKIMGAGQVASSERGRMTGLVDL  
PHNPDELRRLTAVILAGSEDOVAVRADAPRARRLSPAHYATSEVAVPGTLLVYTG  
TAGIAGVARMILMGRGAHEHLAVSRGPDTEGVDLTAELTRLAGAVSAVCSVSE  
PARELVHGLIEQDVVGVVAAAGLPOOVAIINDMEBAPEDEVAAKGAHVHDLCS  
DAELFILFSSGAGVWGSAHQGAAYAGNAPLDAFAHRHGRGLPATYAMKLTAAACRP  
GDEBAVSFLAERGVRAMPVRALAADRVLASGTAVVTVTDVMPABEYTAAPRP  
LIDRTVTTAPSERGAPETSLDRDLGLPRARFATLVLVSTSTAVYGHDDPKAV  
RATTPPEKELGFDLSLAVRNLNLAATGLPSTLVDPHNSAVAGFDLAEISSEVR  
GEAPSLAIGDALAEALPEVPATERBELVQRLRMLAALPRVAAADSGTANPSGD  
DLCEAGVDLELLAIGRELDGNSHHHHH"  
10815. 10861  
terminator  
/note="T7 terminator"  
10882. 10899  
promoter  
/note="T7 promoter"  
10902. 10926  
/note="lac operator"  
10933. 10938  
/note="eryAII (morphed)"  
10933. 10938  
/gene="eryAII (morphed)"  
10947. 10967  
/gene="eryAII (morphed)"  
/note="eryAII (morphed)"  
/note="similar to Saccharopolyspora erythraea EryAII of  
GenBank Accession Number A166156; contains His-tag"  
/codon\_start=1  
/transl\_table=1  
/product="EryAII (morphed)"  
/protein\_id="AAV39551.1"  
/db\_xref="GI:54778586"  
/translation="MTDSEKVAEYLRATLIDLPAAQRILREHSDPIAYSMACRLPG  
GVNTPQRLWELLREGGSTLSEPTDGRGMDLARIHPDPNGTSYVDKGGFLDDAGF  
DAEFPVSPREAAAMPQOQLILREIYENAGIDIPHSILGATVGLFVAGFSGE  
DPAADVEGYSVTGVAFAVASGRISYTMELGSPISVDPAACSSVLAHLAVESLRK  
GESSMAVGGAAVMATPGVDFGFSRORALADGRSKRAGPAGPSSBVTYLLER  
LSBRNGHGVLAVMGSAINODGANGSAPSPRORIRIOTALSCSCEPDDVAV  
EAGTGTALADPTLEANAUDTGRDADLPWLSVKSITGTOAAAGTGLAKVL  
ALNGBLPTALHVEBTPPHVMSGGVALLAQNPWRGERTRAVSAFSGTNAH  
VIVERABERHRETTAHDGRVPLVVSASTALRAQAQIAELRLRPADLIGVGLG  
LATTRARHEHRAVVASTEAEVAGLREIAAGATADAVEGTVRVDGRVPLPQO  
GSONMAGMGLSSSPVPAKIRACDESNAMPQDMVSDVLRQAPAGLDRDVDP  
VLPAVNVSLAELMRSYGVPRAAVNGHSCOEIAAHYAGLTLEDAAKVYGRSLRS  
LSGGGMAVALAEBAVREKLPWQDRLAVNGRPSVSVSEPALRSGDSELCAB  
YWRNLRRETVRFADAVTRLAESGYDAFIEVSPHVVOAVEBAVEDAGEDVAVGS  
LHDDGDLAFLSMATVAHSGVDIRMDVALPGAAPLPTYPEKRYWLOPAAVA  
ASDELAVRSMPTIEKDESGNLDGMLVVTPLISPEWTEMLCEA INANGRALRCVD  
TASRTEMAQVAAQCTGFRGVTLLSSDSACRPVPRAGAVGLITLVQALDAGVDA  
PWCILQGAVRTPADDDLPAPQTTAAGFAQVAGLPGKMGVVDLPSVDAALRL  
LVAVLADGRGAEDHLAVRDGRGRVVAASPOSGRSTPGTGLVGLGASPVQDQ  
LYVMLADGRGAEDHLAVLAGACPGDILAAVEBAGASAVCAADAALEALGDEVTALV  
HAGTLTFNPSIIEVAPEEPAETIAAKTALLAVLDVLAGRAVREYVCSGVATWGA  
GMAAVAGSAVLDALABHRAERGRCTSVATPMAALPGAUVDTYLRBGRSLSDR  
AKRTERVYLAAGVSVAVADVDPVTSSEGPATRPALPHELAGRGAGQABEDSGPT  
GEPAQRLAGLSPDEQENLLELVANAVAEVLGHSAAEINVRFAFSLGIDSTANAL  
RRRLASTGTRLPASLVFDPHTVLAQHLRALSSDAQAAARVAGGADSEPIAIV  
GIGCRPGGIGSPBOLMRYLAEGANLTGFPADRGMIDIGLHPDNDGTSYVDVGG  
FLTDADPDGPGFGITPRBALAMPQOQLILREIYENAGIDIPHSILGATVGLFV  
MNGOSYMOILAGABERYDQYCGLSASVLSGRIATTFEGEGALVYDRACSSLVGI  
HLMQALRREBGSALAGVTVMSDYTTTDFSTOGSLSDGCKFSABADPALS  
GVAALVLEPLSRANGQVLAIRGSANVODASNGLAIPGDSERYIROLAASA  
VPAADVVAEAGHGTGLBPIEAGALITYGQDRPRLTSGVKNIGTQOAAAGA  
GVILVLAARHGMLPSRLADELSPIIDGESCAVEYREVPACERPAACVSGFV  
VSGTNAHVIVEBAPEAOEAPVTERGFLPVLISGRSAVVAAOARALAEHLRDELGL  
TAAATLTAAGRAEPDVAIVTDDBRAGVCAEDLALAGRSADAVA PYMSAPKPYL  
PRGGAGQWGMARDLLESSEVPABESRCAEALSPPTDKLDDVYGGDPPHEHVD  
VLQPVFSTMVSLAEIMRAHGVTPAAVGHSGEIAAAVAGALSLAALKVVALSQ  
VIRELDDQGGSVGASRDELFTVLAARMQGVAAVAVNGGTSVAGPAAEIDEPFAE  
AABREKPRRIARVASHSPVARIEDRLAABIGITTAVAGSPHLSITVGEYIDTSA  
MASVYWRNLRPVLEQAVRGLVEGQFDTFVEVSPHVLVLAAMESTALBAHAEVTVCV

PTLRBQSGPHEBRLNLRHAYHGVGADLRPAVAGRPARELPTYPEHGRFMRPHRP  
ADVALAGVNGAHEHRLVLAADVPRGGAAPFTGRLSTDGOWMLAEHVYGGTLYVGSVL  
VDVLAAGVDGLPVLEELVDPREVLAGAGALRMSVCA PDSEGRATIDVHAEDVLA  
DLADAPWSQHATTTLAGTGAAGRPDTEQPPEDAVRI PLDDHTDGLABQCTYGPSTQ  
ALPAARKDQSVVAEVSIAADEGVAPHTVPAGHLVAVQTSLSLAGEGCKPLPAMNT  
VTLHAGSASVVRVAVPAGADMAALVETPAGHLVAVTVSLVVRSTCKEKOEPERCG  
GEBLHMDWRLAPESTGRVVAADSDIDA VLRSEPEBDAVLVRYBPEDPRRAA  
RGOVLAAALVVRMLBOEELPGA TLYATSGAVTVSDOS VPERGAAMGVIRCAQA  
ESPRVVLDPBAEPMLAVPNDPQALARGDVPVPRLSPLAPSLTLPACGTORLYP  
GDSALDSVAPPEAPVDQYPRAGEVADVRAATVNRDVILALGMPQKMDGVLGALG  
VTVAVGPVDVAPAGDRVIGLFGAAPAPLAVTDHRLIARVPDMSDAVAAVPAAYT  
AHYALHDLGRLAGQSVLTHAAGGVMAVALARAGAEVLATAGARKEGTALALG  
DDEHILSSSETGRAPKFRRTGRGVVDVLSLTGSEBDAVLADGDFVEMGKTD  
LRADGPFGRVAPFDIGRADDBELGELREVVGLLAGEVLDLPVSMELTGA PALO  
HMSGRHVKVLTOPAPVDPDQTVITGCTVLSGLLRLHILYTHRCVRLILVSRG  
ADAPGDELPABIEDJASAEIACOTDRBDSALLDGLRFLRTGVHAGVLADEL  
VTSIDEPAVEQVIRAKVADAMNHLHTANTGSLPVLFSASAVLAPGGGVAAANE  
SLNALAALRTTRGLPAKALGMQASMTSGIDRIRARTVVALPTERBALFDSA  
LRRGSEVPEPLSINRSALRRAEVPEVLKMGVAKLRAQAQAAGPNNVDRLAGRSB  
SDQVAGLAEVLSHAAVSGYSGADQLPERKAKQDGFPSLAVELRNRGTATNGVRL  
PSTLVDPHPTPLA VAEHLDRLSASAPVADICGRULDELEKALEALSNDGHDVQDL  
ESLRRMNSRRADPSTSAISBDASDELFSMLDQRFGGGBDLSGNSHHHHH"  
21742. 21768  
terminator  
/note="T7 terminator"  
21809. 21826  
promoter  
/note="T7 promoter"  
21829. 21853  
misc\_feature  
/note="lac operator"  
21860. 31407  
gene  
/gene="eryAII (morphed)"  
Query Match 46.8%; Score 281; DB 11; Length 31531;  
Best Local Similarity 66.7%; Freq. No. 5.3e-20;  
Matches 401; Conservative 0; Mismatches 200; Indels 0; Gaps 0;  
Qy 1 GCCTCTAGCGCGCGGCAACGCCTTCTCGACGCGCTTGGCGGTGAGCGAGT 60  
Db 2651 GCCTCTAGCGCGCGTGGCAACGGCTTCTTGAATGCTTTCGACGCGCGGAGATCTGAA 25710  
Qy 61 GGGCTCGCGATCACTCGATCGCTCGGAGTCTGTGGAGCGGAGCAATGCGCGTACC 120  
Db 25711 GGGCGACCGGTCAAGATTCGATAGGGAGGTGTGGCGGTGAGCAATGCGCGGTGAT 25770  
Qy 121 GAGGGGGGAGCTACCTGGCGGAGCGGCGCTGGCCATGGAACCGGACGGGGGATC 180  
Db 25771 GAAAGCGGTGATCTGTGTGAGCGGCGGCGGCGGCAATGGAACCAATCTGTGCGGTG 25830  
Qy 181 GAGGAGCTCGGAGCAACCTGGAACGCGGCGGAGCGGCGGTGTGCGGTGAGCACTGAGC 240  
Db 25831 GAAAGACTGCAATTCAGCTGATTCAGCGGTGAGCAACCTCGGTCTCAAGTGTGATATGAC 25890  
Qy 241 GCGGAGCGGTGTGCGAATCTGTTCAACGCGCGCGCGCGCGCGCTCTTTCAGCAATTC 300  
Db 25891 CGTCGCGTGTGTGAGTGTGTTCAACGCGGTGCGCGGTGAGCAACCTCTTGTGTTGATGAATC 25950  
Qy 301 GGTGGGGTCTCGCGCGGCGGCGGAGAGACCGGTCAAGAAATCGAATTCGCGCGGCGGCTG 360  
Db 25951 GCGGGGTGCAACGGGCGGAACTCGCGAGATGAAAGAGGAGCGGTGCGCGGCGGCTCTG 26010  
Qy 361 GCGTGAATCCGGAAGCGGAACGTGACAGGCAATGTCGCGCGGTGTGCGAGCGGAGGTC 420  
Db 26011 GCGGCACTGTCTAACCGCGGAGCGCGCGGAGCACTGCGCAACTGATCCGTGCGGAAATG 26070  
Qy 421 GCAAGCGGTCTGCGGCGCAACGCGCAACGAGTATCGAGCGTGAACGTGCGCTTTCGCTGAC 480  
Db 26071 GCAAGCGGTCTTGTGATCAACGCGGCGAGTGGCGGATGGAACGAGATTCGTGATTCGCGAT 26130  
Qy 481 CTGGGATTCGATTCATACCGCGGTGACCTGCGGAAACCGGCTGCGCGGCGGTGACCGGG 540  
Db 26131 CTGGGGTTTGATCTCAATGACTCGCTGACCTGCGGAAACCGCTTTCGCAACCGGTGACGGGG 26190  
Qy 541 GTCCGGGTGGCAACGATCTGCTTTCGACCAACCGGAGAGTGGAGCGGCTTCACCGGCGAC 600

Db 26191 GTAGGTGAGGCTGCACAGTTGATTGACCATCAACGATCACGCGCTGGCGGATCAT 26250  
QY 601 T 601  
Db 26251 T 26251

RESULT 10  
AY623658/c  
LOCUS DEFINITION  
AY623658 61845 bp DNA linear BCT 16-FEB-2005  
Aeromicrobium erythreum putative transcriptional repressor and  
putative dehydrogenase/reductase genes, complete cds; erythromycin  
biosynthesis gene cluster, complete sequence; putative  
oxidoreductase and lipid genes, complete cds; monoamine oxidase  
gene, partial cds; and unknown gene.  
AY623658  
AY623658.2 GI:59723038

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Aeromicrobium erythreum  
Aeromicrobium erythreum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Propionibacteriaceae; Nocardioidaceae; Aeromicrobium.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
15557441  
2 (bases 1 to 61845)  
Brikun,I.A., Reeves,A.R. and Weber,J.M.  
Direct Submission  
Submitted (11-MAY-2004) Fernalogic, Inc., W. Campbell Park Drive,  
Chicago, IL 60612, USA  
3 (bases 1 to 61845)  
Brikun,I.A., Reeves,A.R. and Weber,J.M.  
Direct Submission  
Submitted (14-FEB-2005) Fernalogic, Inc., W. Campbell Park Drive,  
Chicago, IL 60612, USA  
Sequence update by submitter  
On Feb 14, 2005 this sequence version replaced gi:53794556.  
Location/Qualifiers  
1. .61845  
/organism="Aeromicrobium erythreum"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:2041"  
complement(228. .869)  
/note="ORF13; cecr family"  
/codon\_start=1  
/transl\_table=1  
/product="putative transcriptional repressor"  
/protein\_id="AAU93784.2"  
/db\_xref="GI:59723038"  
/translation="MISAPGESEVSLRAVCAAGVQMPTLYHFGNKGQLTDAYVERG  
FDLYAEKSAGSTGDPIDIRDMDAVAGLNPFGTYTLGKVRGSPSAQAR  
SELRLRLKRAHQRLCTPEQAANHLVNLGVTLRQIVLATEDRDLSTAVRGVI  
AATGSASSPSLQADVRLVLEHANTQRLDGGRETLTKMLSLAVGSDG"  
925. .1746  
/note="ORF12; dehydrogenase/reductase fabG family"  
/codon\_start=1  
/transl\_table=1  
/product="putative dehydrogenase/reductase"  
/protein\_id="AAU93785.1"  
/db\_xref="GI:53794558"  
/translation="MSLOYSCIADPTCSLIDIEGYMSTHTLSEKTVLVAGAKNLGG  
LISROAAGANVALHTNSASTRADAEETLAAVENAGKGVILTGDLVFNVAALFA  
DAESALGVDAVAVTVGVLRKPLADITBAYDAMFDINSKAAVFFIIEARHVAAG  
KVITIVSLGAFDTGYSVAGGSKVHFHFAAKAEERGI SVTALAPSPMDTPFF  
YGGTPEVERVHRSGAMGQLTQIEDISPIRVPLATEGMWITGQITFANGVYTR"  
2172. .2729  
/note="ORF11"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"

CDS

gene  
CDS  
/protein\_id="AAU93786.2"  
/db\_xref="GI:59723040"  
/translation="MPSASRRGRRLTLADCDSPAOPENPACLPWHPPSPSTRRTNP  
OSRPASGORPATPESSSSSRPTSRVAATGAGVALGLTRSLIEGHPTTSGAAAO  
LSTSEIGCPKATGEQREHQAQAPDAOTPRPDRPCKRTPTPIYRRPARRAGADDEP  
GIRRRCTLATLRDAVDPRGDLAG"  
2920. .3813  
/gene="eryD1"  
2920. .3813  
/gene="eryC1"  
3822. .4937  
/gene="eryC1"  
3822. .4937  
/gene="eryC1"  
/codon\_start=1  
/transl\_table=1  
/product="trpD-glucose synthase"  
/protein\_id="AAU93787.1"  
/db\_xref="GI:53794560"  
/translation="MRGILLAGGGRSLHPITRGISKQVLYVDKPMVYPLSTMLA  
GIRDILITSPHEALFVRLIDGSAAGVNLSTAVOPTPGDLAFLIGSHLQGEA  
ALVLGDNITFHGTMGVQLRRHTQTEVAATIGRTSNFOAGVVEIGEDRALSLERK  
QRPSPYAVPGLFYDPDVVEHAKTLVPSARGLEITDLRLYLQGRLLVVEVLPGT  
AMFDGTGFDLSLNDASNFVRTIARQGLVGSPEVAVRVGFLSDABILLERSSST  
YGYLRALDREGDADHPYRV"  
3822. .4937  
/gene="eryC1"  
3822. .4937  
/gene="eryC1"  
/codon\_start=1  
/transl\_table=1  
/product="transaminase"  
/protein\_id="AAU93788.2"  
/db\_xref="GI:59723041"  
/translation="MTIDVPLDIDAAVEIRPOIDAAVSGVLAAGRYLLAQTEARE  
AEFAAASAHCVTSGCDALSELSTLALGVDPQDEVVAVPFIATMLVSRGAP  
VPSPEPDGLVDEVAEAAITPRTALILVHLXGEADALAIRVDRRLALVEA  
AOSTGARGDGAAGVSGSTAAFSFYGKQILAGDGAAYTADDDLARVRLRNG  
SVAKYHVRVGTNSRLDIEQAIVLRATLPLIDMNAERATIAIYLRLEASGVRLP  
PHDGRSAHMLFVVRCHDRETLQBELTRHGVETLVHPTFVHLSAAYADHGRAGSLP  
HAERLARVSLTLPWGPHLPPLAERVADLVALAAR"  
/complement(4934. .5527)  
/gene="eryB11"  
/complement(4934. .5527)  
/gene="eryB11"  
/codon\_start=1  
/transl\_table=1  
/product="trpD-4-keo-6-deoxyglucose 3,5-epimerase"  
/protein\_id="AAU93789.1"  
/db\_xref="GI:53794562"  
/translation="MRAQRVAGGMEFTPTVHNDAGVPASPIQEEAFVAAVGHPP  
VAQTNHLVSDRGVLRGVHFTASPEQSKTVHCAARALDMLDRTDSFTOTMAVEL  
DPRCNALYFPIGTAAAFMALPEQTINSYLVSTYDASLBIALIDPFDPELALPMPADV  
PALLSGRDEVAMSPAQARAGLLPRSGPRVABGEVEP"  
/complement(5530. .6969)  
/gene="eryCV"  
/complement(5530. .6969)  
/gene="eryCV"  
/codon\_start=1  
/transl\_table=1  
/product="deoxyhexose reductase"  
/protein\_id="AAU93790.1"  
/db\_xref="GI:53794563"  
/translation="METSMMKLVDAVARGVDHHTAKRDVLAERLRSKMGSDGVAKAAHLA  
ALVSHYAHPEPTPLEQROKRGIDRDSFAALVAFDRPELTALAVERGPAKYASNTI  
KPLNEAGALDAVERRAPFVSVGLYKPTCMFCHCVRTGARVYADSLTPNEVLA  
ASVLEVDNDORTTMYSGGLEPLNPGLEGILIRANRGGLDMRTYNSFALTATLS  
ROGVNPTGLALRTSLVGLNDEEVEETTKGA FSRVAVLORFMHGOEBSRDLIG  
NHITLPRGTGLILVDPLIEALNSSDRPLDPTVYEDVSGRDGLDRSREBILH  
GLDFMEYAEKHTHGMITIDGYALBSLRGVDAELKIRPTMRPLAHQVAAVOVDL  
GDVLYEAGAPDDIDAGRRYIAGVTSDDSHLDVLRFEVSGREVBAPQDSYFLDGP  
DQSVATLNLQDLDVADQWGRFGLVNAS"  
/complement(6997. .8202)  
/gene="eryCIV"  
/complement(6997. .8202)  
/gene="eryCIV"  
/codon\_start=1

Query Match	Score	DB 1	Length
Beet Local Similarity	66.3%	Pred. No. 2,6e-18;	
Matches 396;	Conservative 0;	Mismatches 195;	Indels 6; Gaps 1;
5	CCTAACGGGGGCAACGCGCTTCTCCAGCGCTTCCGCGCGTTCGAGCGGAGGAGGGG	64	
Db	29335	CCTAACGGCGCGCAACGCGCTTCTCCAGCGGCGTTCGCGCGAGCGGCGGACCGG	64
		29278	
	65	TGCGGTCACCTCGATCGCCTGGGGCTGTGGGCGGGGCAAACTAGCCGATACCGAGG	124

REFERENCE	ORGANISM	KEYWORDS	VERSION	ACCESSION	LOCUS	DEFINITION	AB088224/c	RESULT 11
Db	29277	---	CGTACCTTCGTCGCGCTGAGGCGCGGACGACCAATGCGCGCGCTCGAGG	29222				
Qy	125	GCAGCGCATCTACCTTGCAGCGCAAGGCGCTGCGCGCATATGACACCGGACGCGGCGATCGAGG	184					
Db	29221	GCAACACCTTACCTTGAAGGCGCAAGGCGCTGCGCGCATATGCGCGCGGACGCGATGCGG	29162					
Qy	185	AGCTGCGGACCACTTGAAGCGCGGCGGACCGCTGAGGTGTCTGATGTGTGAACCTTGAACCGG	244					
Db	29161	GCTGTGAGAGACCGCGCTTCAACCGGAGGAGACACTGCGCTTCTCGTGTGCACTTGAACCTG	29102					
Qy	245	AGCGGTTCTGTGAACCTTTCACCGCGCGCGCGCGCGCGCGCGCTTCTTTCGACGAACTTGGTG	304					
Db	29101	ACGCTTCTGTGACGATGTCTTTCGCGCGCGGACGCGCTTCTTCTTTCGACCGGCTGCGC	29042					
Qy	305	GAGTCCGCGCGCGGCGGAGACCGGATGAGGATTCGATTCGCGCGCGCGCGCGCTG	364					
Db	29041	CCGCGCGCGCGGAGGCG	28982					
Qy	365	CGATGCGGAGGCGCGGACGTCACGACGATGTGCGCGCGCTGCTGCGAGCGGATGCGAG	424					
Db	28981	CCAAGACGCGCGGAGGAGCGCGCGCTGTCTGAGGTCGTGTGCAACCGCGAGGATGTG	28922					
Qy	425	CGGTGCTGCTG	484					
Db	28921	GTTGCTCTGCGCGCTGACGACGACGCGCGCGTGAACCGGACGCTTCTTCTCGCGACTTG	28862					
Qy	485	GATTGCACTGCATGACCGCGCGCTGACCTGTGAGGACCGGCTGCGCGCGCGGACCGGCGCTC	544					
Db	28861	GATTGCACTGCATGACCGCGCGCTGACCTGTGAGGACCGGCTGCGCGCGCGGACCGGCGCTC	28802					
Qy	545	GAGTGGCGACGACCATCTTCTTTCGACCAACCGGACGATGAGCGCGCTTCAACCGCGCACT	601					
Db	28801	GCGAGGCGCGGACCGGATGTGCTTTCGACCAACCGGCTTCAACCGGACGATGAGCGCGCACT	28745					
REFERENCE	11	AB088224	210614 bp	DNA	linear	BCT 11-JUN-2003		
LOCUS	AB088224/c							
DEFINITION	Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.							
ACCESSION	AB088224							
VERSION	AB088224.1	GI:30698345						
KEYWORDS	Streptomyces rochei							
SOURCE	Streptomyces rochei							
ORGANISM	Streptomyces rochei							
REFERENCE	1	Kinaeshi, H., Fujii, S., Hatani, A., Kurokawa, T., and Shinkawa, H.						
AUTHORS	1	Physical mapping of the linear plasmid pSLA2-L and localization of						
TITLE	1	the eyaI and actI homologs						
JOURNAL	2	Bioest. Biotech. Biochem. 62, 1892-1897 (1998)						
AUTHORS	2	Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H.,						
TITLE	2	Nimi, O., and Kinaeshi, H.						
JOURNAL	3	Identification of two polypeptide synthase gene clusters on the						
PUBMED	3	linear plasmid pSLA2-L in Streptomyces rochei						
TITLE	3	Gene 246 (1-2), 123-131 (2000)						
REFERENCE	4	Hiratsu, K., Mochizuki, S., and Kinaeshi, H.						
AUTHORS	4	Cloning and analysis of the replication origin and the telomeres of						
TITLE	4	the large linear plasmid pSLA2-L in Streptomyces rochei						
JOURNAL	5	Mol. Gen. Genet. 263 (6), 1015-1021 (2000)						
PUBMED	5	10954087						
REFERENCE	6	Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K.						
AUTHORS	6	and Kinaeshi, H.						
TITLE	6	The large linear plasmid pSLA2-L of Streptomyces rochei has an						
JOURNAL	7	unusually condensed gene organization for secondary metabolism						
PUBMED	7	Mol. Microbiol. 48 (6), 1501-1510 (2003)						
REFERENCE	8	12791134						
TITLE	8	(bases 1 to 210614)						

**AUTHORS** Mochizuki,S., Hiraetsu,K. and Kinashi,H.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (15-JUL-2002) Harryasu Kinashi, Hiroshima University.  
Department of Molecular Biotechnology, Graduate School of Advanced  
Sciences of Matter, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima  
739-8530, Japan (E-mail:kinashi@hiroshima-u.ac.jp).  
Tel.:81-824-24-7869, Fax:81-824-24-7869)  
The nucleotide sequence has been determined by using restriction  
fragments and nested deletion fragments of the ordered cosmid  
library of pSLA2-L.  
pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries  
143 ORFs.  
Gene prediction was based on the unique codon usage in Streptomyces  
(Bibb et al., Gene 30:157-66 (1984)) using the Framaplot program of  
Ishikawa and Hotta (FEBS Microbiol Lett 174: 251-253 (1999)) as  
implemented at  
http://www.nih.go.jp/~junc/cgi-bin/framaplot.pl. Where possible we  
chose an initiation codon (atg, gtg, ctg or tgg) which is preceded  
by an upstream ribosome binding site sequence (optimally 5-13 bp  
before the initiation codon). If this could not be identified we  
chose the most upstream initiation codon.  
Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes  
(three PKS gene clusters for lankacidin, lankamycin and an unknown  
type II polyketide, a carotenoid biosynthetic gene cluster, many  
regulatory genes and others). The range of each biosynthetic gene  
cluster has been deduced by comparison with similar gene clusters  
in most cases and may be revised in future.

**FEATURES**  
**SOURCE**  
Location/Qualifiers  
1. .210614  
/organism="Streptomyces rochei"  
/mol\_type="genomic DNA"  
/strain="7434AN4"  
/db\_xref="taxon:1928"  
/plasmid="pSLA2-L"  
/note="linear plasmid"  
1. .1992  
/note="left terminal inverted repeat, TIR-L; shows 99.4 %  
(1981/1992) sequence identity to TIR-R (complement  
(208623. .210614))"  
683. .2188  
/note="N-terminal sequence is almost identical (435/437)  
with that of ORF143 at the right end of pSLA2-L until the  
inner end of TIR-L  
ORF1 (501 aa)  
similar to A1590463 Streptomyces coelicolor putative  
helicase, SCPI.136 (879 aa); homology is seen until the  
inner end of TIR-L"  
/codon\_start=1  
/transl\_table=1  
/product="putative helicase"  
/protein\_id="BAC76459.1"  
/db\_xref="GI:30698346"  
/translacion="MSTTSRTDQREBAQDEAVDAVVRALIEPVLAPERGLRQVIM  
ATGSGKTRVARSAEKLAKGRLVLPBLDTQTEAMRAGRTGWTIGVSLRGED  
VAPNTDVEELVWVVPFDKVTVPATASIGLTERAHRGGLPGMDLIVDAHRT  
SGRLGKPAVAVDNDTRILPSRLVMTATPRLMOLDEDEBAGELVASMEDPDGLG  
ARCTTITSEALIDGICAPVOVCVDITDLOAOLLGVRSGSDEVRGARTLALQTA  
LTKASSENPRRTLVFHHMYKEARPAAGLDVAKRLAAAPGLYPRITVANNICGHI  
KPGHRRVIGSFSTGIATDGTVEKEGFLSGVKVAGSDTRCESVWADRGSPDL  
VQVAGRALRIPGQGVKASLVVPLLDGETVADNMLTSPYNGALLERQAPLLTGHG  
SEGFRAALPGFEVAPVGVRAQGRGASGASRSITPSAMSSISISWRAP"  
complement (2315. .3595)  
/note="ORF2 (426 aa)  
similar to A8004736-10 Pseudomonas aeruginosa hypothetical  
protein (442 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="putative secreted protein"  
/protein\_id="BAC76460.1"  
/db\_xref="GI:30698347"  
/translacion="MRRKPRALLGLPIGLLVAVAPADAHAASADTVGTWAPPTAA  
PAADSTVEDQTLQTVHLSVAGDSLRTFRFNEPSTPLTIGEVAAAPAAQGPATV  
DPGIDRVVRFGGRPSATLAPGQWSDVVALPTTAPGDLVLSLTPRTPGSTVSHAA

**CDS**  
1  
/transl\_table=1  
/product="Str-like regulatory protein"  
/protein\_id="BAC76461.1"  
/db\_xref="GI:30698348"  
/translacion="MTAALAGNIDQETRSIPISLVPGSPRLCGSDAEHVARLADTD  
BALPILVERSTRTIDGMHRYLAKAKNGTTEVRLFDQAASAFLAIVRSNTHGL  
PLSRORRAAAGRTILIAOWPHLSRBAVAGICAKTYVALRPAAGTTPAPKARGD  
GRTRPLDGTGRRKABLQRPQASVREYARHAGISPATASVYRRRLAAGRPVPR  
NTPGARPAAGTGRATPGGAAPSVSVRPIRPEDSPVRLRLDLSHKSGRLL  
LRLQCAVARTLALMAQVTPPHCTDVLVELAREYADLWAEFARVARTDG"  
complement (4922. .5812)  
/note="ORF4 (296 aa), lankacidin biosynthesis protein  
similar to A8004625-4 Pseudomonas aeruginosa  
pyrroloquinoline quinone biosynthesis protein B (304 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="pyrroloquinoline quinone biosynthesis protein B"  
/protein\_id="BAC76462.1"  
/db\_xref="GI:30698349"  
/translacion="MILLSTAGGGFPQNNCAALCARGRGELPARSGECVAVSGDG  
RDWMLNASPDIRIOLLAAPALTPGPBPRTPPVAGVLTATBAVDHALGLVLRATGL  
TVYAAPVPGRALSAELPVRGLIDRYAPMDRDAAPGFAVAGGLTVYAPVGTAKK  
YAAPDPDAPVCARTRIEDPATGALVYACLTATWPGFDLISATCALDGFPA  
GELGATSSAGAGOSLMGHLPVAGPGSLAALAHRLRRTYTHLNTNPLDPSAA  
HAAREAGVEVLPDGSLEY"  
complement (5815. .6909)  
/note="ORF5 (364 aa), lankacidin biosynthesis protein  
similar to A1277117-5 Gluconobacter oxydans putative pqg  
protein (359 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="pyrroloquinoline quinone biosynthesis protein E"  
/protein\_id="BAC76463.1"  
/db\_xref="GI:30698350"  
/translacion="MADPAGAPACMLIELTRCPHCPYCSNPLEYRREAEYCEQ  
WTDILNARIELGVOMHFGSGEPRLARDLDIVGHARLIGAVVLLVSGVGLTBRH  
DLARQVDYVQLSIQDDPRAAGAIACARHTATLGAAYTAAGLPLTVYIVYHRN  
IDRTGRVVDLAVDLAGDRIELANTQYIGKGRNRAALPTAQLAAAREAVRARTRI  
AGGELVYVADVYDDRPKPCMDGMSITQILVPAGVLPAPAAVATTTLPVSNALRR  
PLSEIYASRSFNAYRGTMWRBECRCPSRHADHGGCRCAFQLTDDAATDPAQGL  
SPHRSIVDALAEVTDGVPVAFVPRGVPA"  
complement (6902. .7174)  
/note="ORF6 (90 aa), lankacidin biosynthesis protein  
similar to A603642-200 Shirotozodium melioli putative  
pyrroloquinoline quinone biosynthesis protein D (98 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="pyrroloquinoline quinone biosynthesis protein D"  
/protein\_id="BAC76464.1"  
/db\_xref="GI:30698351"  
/translacion="MTGLPEPTVRLRPGVRLTRDPARGSLALLPERVVVLTNDTAAV  
LAHCDGTTSLAGIYRLAEYGVASADVEIELRLAQRRVVDLHG"  
complement (7171. .7899)  
/note="ORF7 (242 aa), lankacidin biosynthesis protein  
similar to A8004625-5 Pseudomonas aeruginosa  
pyrroloquinoline quinone biosynthesis protein C (250 aa)"  
/codon\_start=1  
/transl\_table=1  
/product="pyrroloquinoline quinone biosynthesis protein C"  
/protein\_id="BAC76465.1"  
/db\_xref="GI:30698352"  
/translacion="MSMSVTRKVAAPMSEAFRQRLALASSYWDRRPFRHMHGGL  
DEGLRLMAANRWYORCLPKQDAIIVANCLPEVRKQMSRIYVHIDGADACAGAEK



## CDS

WLRLAENVGLRDEVDRLVLATGTPAVADAYVDFARRRPMLEAAAGLTSPGLL  
AHRIGLRREHPWIAEGFEYFARIEIVGPEKRSUDDLVARRAVSRBQEQACVRLAA  
FKGRVNAVLDSDHYHNGATRS"  
complement(7926..8030)  
/note="ORF8 (34 aa), lankacidin biosynthesis protein  
similar to AL603642-197 Sinorhizobium meliloti putative  
pyrroloquinoline quinone biosynthesis protein A (31 aa)"  
/codon\_start=1  
/transl\_table=11  
/product="pyrroloquinoline quinone biosynthesis protein A"  
/protein\_id="BAC76466.1"  
/db\_xref="GI:30698353"  
/translation="MRTSGKELPAKKAMHRPDVTITDMEVATYFSR"  
8420..10057  
/note="ORF9 (545 aa), possible lankacidin resistance  
protein  
similar to AL133424-19 Streptomyces coelicolor putative  
ABC transporter ATP-binding protein (544 aa)"  
/codon\_start=1

## CDS

Query Match 40.5%; Score 243.6; DB 1; Length 210614;  
Best Local Similarity 65.1%; Pred. No. 1.5e-16;  
Matches 401; Conservative 0; Mismatches 194; Indels 21; Gaps 2;  
QY 1 GCCTCTACGCGCGGCGCAAGCCCTTCTCTGACGCTTCCGCGTGGTGGCGAGT 60  
Db 60245 GGGCGCTACCGCGCGGCGCAAGCCCTTCTCTGACGCTTCCGCGCGCGCGCGC 60186  
QY 61 GGGCTGCGCGTCACTTCGATCGCTGGGATCTGTGGCGCGGCGCAACATGGCGGTAAC 120  
Db 60185 GGGCTGCGCGTCACTTCGATCGCTGGGATCTGTGGCGCGGCGCAACATGGCGGTAAC 60126  
QY 121 GAGGCGGCGCACTACCTGCGCGCAAGCGGCGCTGCGCGCATGACCGCGCGCGATC 180  
Db 60125 GCGGCGCGCACTACCTGCGCGCAAGCGGCGCTGCGCGCATGACCGCGCGCGATC 60066  
QY 181 GAGGAGTGGCGCAAGCTTGGAGCGCGGCGGACCGGCGGATCGAGTGGAGCTGGAG 240  
Db 60065 GCGGAGTGGCGCAAGCTTGGAGCGCGGCGGACCGGCGGATCGAGTGGAGCTGGAG 60006  
QY 241 CGGAGCGGATTCGTGCACTGTTCAACCGCGCGCGCGCGCGCGCTTCTTCAAGAACTC 300  
Db 60005 CGGAGCGGATTCGTGCACTGTTCAACCGCGCGCGCGCGCGCGCTTCTTCAAGAACTC 59946  
QY 301 GGTGGGATCG 351  
Db 59945 GGTGGGATCG 59886  
QY 352 CGGCGGCTGG-----CGTGAATGCGGAGCGCGAGCTGACGAGCAATGTCGCC 399  
Db 59885 TCCCGGCTGGCGGAGCTGCTGCGCTGCGCGCGCGGAGCGCGCGGAGCACTGCTG 59826  
QY 400 CGGCTGTCGAGCGCGAGGTGGAGCGGTGCTGGGCGCAACCGCGCAAGCTGATCGAG 459  
Db 59825 CGGCTGTCGAGCGCGAGGTGGAGCGGTGCTGGGCGCAACCGCGCAAGCTGATCGAG 59766  
QY 460 CGTGAAGTGGCTTCTGCTGACCTTGGAGATTCATGACCGCGCTGCACTTGGGAGAC 519  
Db 59765 CGTGAAGTGGCTTCTGCTGACCTTGGAGATTCATGACCGCGCTGCACTTGGGAGAC 59706  
QY 520 CGGCTGCGCGCGGTGACCGGCGGTGCGGAGTGGCGAGCAATGCTTTCGACCGCGACA 579  
Db 59705 CGGCTGCGCGCGGTGACCGGCGGTGCGGAGTGGCGAGCAATGCTTTCGACCGCGACA 59646  
QY 580 GTGAGACCGGCTCAACG 595  
Db 59645 GTGAGACCGGCTCAACG 59630

RESULT 12  
AX697993 4725 bp DNA linear PAT 02-APR-2003  
LOCUS AX697993  
DEFINITION Sequence 17 from Patent WO03010193.

ACCESSION AX697993  
VERSION AX697993.1 GI:29499042  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacterium; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporinae; Micromonosporaceae; Micromonospora.

REFERENCE  
1 Earnest, C.M., Staffa, A. and Yang, X.  
Genes and proteins for the biosynthesis of rosamycin  
Patent: WO 03010193-A 17 06-FEB-2003;  
JOURNAL Biosciences Inc. (CA)  
FEATURES  
source location/Qualifiers  
1..4725  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 36.6%; Score 220.2; DB 6; Length 4725;  
Best Local Similarity 61.6%; Pred. No. 1.2e-13;  
Matches 370; Conservative 0; Mismatches 228; Indels 3; Gaps 1;  
QY 3 CTCCTACGCGCGGCGCAAGCCCTTCTCTGACGCTTCCGCGTGGTGGCGAGTGG 62  
Db 3879 CGGCTACGCGCGGCGCAAGCCCTTCTCTGACGCTTCCGCGCGCGCGCGCGCG 3938  
QY 63 GCTCCGCGTCACTTCGATCGCTGGGATCTGTGGCGCGGCGCAACATGGCGGTAAC 122  
Db 3939 GCTGACCGCGCACTTCGATCGCTGGGATCTGTGGCGCGGCGCAACATGGCGGTAAC 3998  
QY 123 GGGCGCGCACTACCTGCGCGCAAGCGGCGCTGCGCGCATGACCGCGAGGCGCATGCA 182  
Db 3999 GCGCGAGGATACCTGAGCGCGCGGCGCTGCGCGCATGACCTGCGAGGCGCATGCA 4058  
QY 183 GAGGCTGCGGACCAACCTTGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 242  
Db 4059 CGGCTCTGCG 4118  
QY 243 GAGGCGGATTCGTGCACTGTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 302  
Db 4119 GAGGCGGATTCGTGCACTGTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 4178  
QY 303 TGGGATTCG 359  
Db 4179 GAGGCG 4238  
QY 360 GGGCTGATGCGCGGAGCGCGGAGCTGACGAGCAATGTCGCCCGGCTGTCGAGCACT 419  
Db 4239 CGCGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAG 4298  
QY 420 GGGAGCGGATTCGTGGCGCAAGCGCGCAAGCTGATCGAGCTGAGCTTCCGATG 479  
Db 4299 CGCGCGCGCTCTGCGCACCGGCGCGCGGAGCGCGGAGCGCGGAGCGCGGAG 4358  
QY 480 CCGGAGATTCGACTCATGACCGCGCGCTGCACTTGGGAGACCGGCTGCGCGGATGAC 539  
Db 4359 GATCGGATTCGACTCATGACCGCGCGCTGCACTTGGGAGACCGGCTGCGCGGATGAC 4418  
QY 540 GGTCCGCGGAGCGCGAGCAATGCTTTCGACCAACCGGAGGAGCAACCGGCTGCGCGCA 599  
Db 4419 CGTCCGCTGTCGACCGCTGCTTTCGACCAACCGGAGGAGCAACCGGCTGCGCGCA 4478  
QY 600 C 600  
Db 4479 C 4479

RESULT 13  
AX697977 60196 bp DNA linear PAT 02-APR-2003  
LOCUS AX697977  
DEFINITION Sequence 1 from Patent WO03010193.  
ACCESSION AX697977



[illegible]

KEYWORDS	Streptomyces avermitilis
SOURCE	Streptomyces avermitilis
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE	1
AUTHORS	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinoue, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osone, T., Kikuchi, H., Shida, T., Sakaki, Y. and Hattori, M.
TITLE	Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
PIRMBED	11572948
REFERENCE	2 (bases 1 to 104326)
AUTHORS	Ikeda, H.
TITLE	Direct Submission
JOURNAL	Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of Microbial Chemistry, School of Pharmaceutical Sciences, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan (E-mail: ikeda@mc.pharm.kitasato-u.ac.jp, Tel: +81-3-5791-6242, Fax: +81-3-3444-6197)
FEATURES	Location/Qualifiers
source	1..104326
gene	/organism="Streptomyces avermitilis"
gene	/mol_type="genomic DNA"
gene	/db_xref="taxon:33903"
gene	complement(1..1134)
gene	/gene="mcmb"
gene	complement(1..1134)
gene	/gene="mcmb"
gene	/note="methylmalonyl-CoA mutase"
gene	/codon_start=1
gene	/transl_table=1
gene	/protein_id="BAB69185.1"
gene	/db_xref="GI:15823968"
gene	/translation="MTGSRVFLFGSLRDRVPSLAVDPAAPLIVLYOLVAEHHGAPAO RGIVGRGVAVDSIDDRVFLFGSLRDRVPSLAVDPAAPLIVLYOLVAEHHGAPAO RLAAGVKGDFEFLANGPGVFLRPSRLMANFAYCOALEPWCLOSLVCGHILACT GADPALVEAFVAVAGIERYLRAAVTAGIDVRAVRLCLVGGTGGENKAKLQATRR VMTNAAERFGSGRTAERYLYRTPRDGLATLCTEDIEAGTILVDRIERGTLTVL BHGLAGGLGRPHGDRPATRPPHPODSDLDLQRTERLAKLRAMPVQAVDDEALLR LRKAGQDDNLTLYMRRLAARATVGSVDLNRSTWGTGSPAT"
gene	complement(111..1886)
gene	/gene="ppta"
gene	complement(1131..1886)
gene	/gene="ppta"
gene	/codon_start=1
gene	/transl_table=1
gene	/product="phosphopantetheinyl transferase"
gene	/protein_id="BAB69186.1"
gene	/db_xref="GI:15823969"
gene	/translation="MRAHEVPADVPVSLIHPREALVODAVESRRRREPARGRCARRA LASIGVPGPLATPRGAPRMAGVCGATIGAGYRAAVALRAGIYAGIDADPRGAG LPGVLEAVLPAQGPQYKERTAPHPGVWGKLLFAKESGYKAMPATGAPLGFEEB LIDLPVSTFRVRLPTALRAAPBELDLHGRLVRRGILVTVAVDGPDRDTPHAR SPYRTKSAPTGSGADRPGRPTDRAIDRPTA"
gene	complement(1910..2875)
gene	/gene="simk4 homolog"
gene	complement(1910..2875)
gene	/gene="simk4 homolog"
gene	/codon_start=1
gene	/transl_table=1
gene	/product="simk4 homolog"
gene	/protein_id="BAB69187.1"
gene	/db_xref="GI:15823970"
gene	/translation="MSQDRHGSRTALGSGHGRILAVSDIYHRRNRDVIQDIAH GCAEMLIAGDVAVESVADITWLSLIDRRAKTIWAGNIBLWTAPEDPDLIGER RHLVDICRALGVTPEDPYPMGAGQPVAVAPLFLYDVSFLPRGWSYSHLRLLRLA DYADVICTDEFLIHDPDPSRBAWCRARVATRARLDLPDLPTVLVNHPLVAB RVLRRBEALMGCTATADMLRYAAALVGHGILIPITICDVPHEHVESLGVRSE QRYGRSRPMPVQILTPRAPAPAPVPEVVPAPPGHRTSPADRRP"
gene	3213..39771



Db 22981 TGGAGGCGCTTCGCGGAGCGTTACCGCGGCCCGCCCGCCCTCATCGACGCCATC 23040  
QY 301 GGTGGGGTCCGCGCGCGGGGCC-----GAGGAGCCGGTCAGATCGATCGCCCG 354  
Db 23041 CCGAGGCG 23100  
QY 355 CGGCTGGCGTTCGATTCGCGGAGCGGAACTGTCAGACATGTCGCGCGTGGTCGAGCC 414  
Db 23101 AGGCTGGCGCGGTCGACGCGGAGCGGAACTGTCAGACATGTCGCGCGCGCGCG 23160  
QY 415 GAGGTGGCGAGCGGTCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 474  
Db 23161 GCCCGCGCGCTTCG 23220  
QY 475 CGTACCGTGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 534  
Db 23221 AAGGACTGGGCTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 23280  
QY 535 ACCGCGGTTCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 578  
Db 23281 ACCGCGGTTCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 23324

RESULT 15  
BA000030\_36/c

WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragment Name	Begin	End
BA000030_00	1	110000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000

BA000030\_45 4500001 4610000  
BA000030\_46 4600001 4710000  
BA000030\_47 4700001 4810000  
BA000030\_48 4800001 4910000  
BA000030\_49 4900001 5010000  
BA000030\_50 5000001 5110000  
BA000030\_51 5100001 5210000  
BA000030\_52 5200001 5310000  
BA000030\_53 5300001 5410000  
BA000030\_54 5400001 5510000  
BA000030\_55 5500001 5610000  
BA000030\_56 5600001 5710000  
BA000030\_57 5700001 5810000  
BA000030\_58 5800001 5910000  
BA000030\_59 5900001 6010000  
BA000030\_60 6000001 6110000  
BA000030\_61 6100001 6210000  
BA000030\_62 6200001 6310000  
BA000030\_63 6300001 6410000  
BA000030\_64 6400001 6510000  
BA000030\_65 6500001 6610000  
BA000030\_66 6600001 6710000  
BA000030\_67 6700001 6810000  
BA000030\_68 6800001 6910000  
BA000030\_69 6900001 7010000  
BA000030\_70 7000001 7110000  
BA000030\_71 7100001 7210000  
BA000030\_72 7200001 7310000  
BA000030\_73 7300001 7410000  
BA000030\_74 7400001 7510000  
BA000030\_75 7500001 7610000  
BA000030\_76 7600001 7710000  
BA000030\_77 7700001 7810000  
BA000030\_78 7800001 7910000  
BA000030\_79 7900001 8010000  
BA000030\_80 8000001 8110000  
BA000030\_81 8100001 8210000  
BA000030\_82 8200001 8310000  
BA000030\_83 8300001 8410000  
BA000030\_84 8400001 8510000  
BA000030\_85 8500001 8610000  
BA000030\_86 8600001 8710000  
BA000030\_87 8700001 8810000  
BA000030\_88 8800001 8910000  
BA000030\_89 8900001 9010000  
BA000030\_90 9000001 9100000  
Continuation (37 of 91) of BA000030 from base 3600001 (BA000030 Streptomyces avermitilis)

Query Match 34.9%; Score 210; DB 1; Length 110000;  
Best Local Similarity 61.3%; Pred. No. 4,1e-13;  
Matches 358; Conservative 0; Mismatches 220; Indels 6; Gaps 1;

QY 1 GCTCTTACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60  
Db 15064 GCGCGTACG 15005  
QY 61 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
Db 15004 GGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 14945  
QY 121 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
Db 14944 GAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 14885  
QY 181 GAGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
Db 14884 GCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14825  
QY 241 CGGAGGCGGCTTCGTCGACGTCGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 14824 TGGAGGCGGCTTCGTCGACGTCGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 14765  
QY 301 GGTGGGCTTCG 354

```

Db      14764 CCGAGGCGCGCCCGCCGCGCCAGACGATCCGGCCGCGGAGAGTCCGAACCTGGCCGCC 14705
Qy      355 CCGCTGGCGCTCGATGCCGAGGCCGAACTCAAGCATGTCCGCCGCGCTGGTCCGAGCC 414
Db      14704 AGGCTGGCCCGGATCGACCGAACGGGAAACGCAACTCCTCGACTGGTCAAGGACC 14645
Qy      415 GAGGTGGCAGCGGTGCTGGGCGACCGGCAACGCGGACGCTGATCGAGCGGTGACGTCCCTTC 474
Db      14644 GCCGCCGCCCTCGGCTGGGAGCACCGGGGCGCCGAGGCGATCAACCCGTCGAAAGCCCTTC 14585
Qy      475 CGTACCTGGAGATTGACTCCATGACCGCGTCACTTGGCGAAACGGGCTCGCGCGGTG 534
Db      14584 AAGGACCTGGGCTTCCGACTCCCTGACCGCGCTCGACTTGGCGAAACGGGCTCACCGCCG 14525
Qy      535 ACCGGGGTCCGGGTGGCCAGACCATGCTTTCGACCAACCGAC 578
Db      14524 ACCGGGTACGTCTGCCCGCCAGCTGTCTTCGACCAACCCAC 14481

```

Search completed: April 7, 2006, 16:57:38  
 Job time : 2908.81 secs

**This Page Blank (uspto)**